U.S. Army
Research Institute of
Environmental Medicine

Natick, Massachusetts

TECHNICAL REPORT NO. T07-06
DATE December 2006
ADA 460374

Preliminary Derivation of Test Item Clusters for Predicting Injuries, Poor Physical Performance and Overall Attrition in Basic Combat Training

United States Army
Medical Research & Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 2006	Technical Report

4. TITLE AND SUBTITLE
Preliminary Derivation of Test Item Clusters for Predicting Injuries, Poor Physical Performance, and Overall Attrition in Basic Combat Training

5. FUNDING NUMBERS

6. AUTHOR(S)
Dr. Stephen Allison, Dr. Joseph Knapik, and Ms. Marilyn Sharp

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Military Performance Division
US Army Reserach Institute of Environmental Medicine
42 Kansas Street
Natick, MA 01760-5007

8. PERFORMING ORGANIZATION REPORT NUMBER
T07-06 Allison

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Medical Research and Materiel Command
Fort Detrick
Frederick, MD 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Analytic methods including test item clusters (TICs) employed in medical diagnostic testing have potential for estimating probabilities of negative military training outcomes in individual trainees. Baseline attributes and performance scores that discriminate between groups experiencing negative vs. positive training outcomes were combined to maximize predictive power and accuracy. Predictive models were derived from 15 baseline variables using existing data (518 men and 416 women Basic Combat Training (BCT) trainees) to predict: Army Physical Fitness Test (APFT) failure at week 7; overuse injury during BCT; and failure to complete BCT with peers. The models included from one to four predictors per TIC. Large shifts in pre-test to post-test probability (as high as from 16% pre-test probability to 91% post-test probability) were observed with TICs to predict APFT failure for both men and women, and to predict overuse injuries in men. Smaller probability shifts were seen with the single tests identified to predict BCT attrition for both men and women. No useful model for predicting overuse injuries in women was derived from the methods and data employed in this study. This study suggests good potential for these analytic methods to derive useful combinations of prognostic variables for predicting negative outcomes in BCT.

14. SUBJECT TERMS
Clinical Prediction Rules, Test Item Clusters, Overuse Injury Prediction, APFT Pass/Fail Prediction, Basic Combat Training Attrition, Negative Training Outcomes

15. NUMBER OF PAGES
58

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

DISCLAIMERS

The views, opinions and/or assertions contained in this publication are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

Human subjects participated in these studies after giving their free and informed voluntary consent. Investigators adhered to AR70-25 and USAMRMC Regulation 70-25 on the Use of Volunteers in Research. For protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law CFR 46.

Approved for public release; distribution unlimited.

USARIEM TECHNICAL REPORT T07-06

PRELIMINARY DERIVATION OF TEST ITEM CLUSTERS FOR PREDICTING INJURIES, POOR PHYSICAL PERFORMANCE, AND OVERALL ATTRITION IN BASIC COMBAT TRAINING

Dr. Stephen C. Allison
Dr. Joseph J. Knapik
Ms. Marilyn A. Sharp

December 2006

Military Performance Division
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760-5007

Injury Prevention Program
U.S. Army Center for Health Promotion and Preventive Medicine
Aberdeen Proving Ground, MD 21010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Basic Combat Training (BCT) in the U.S. Army is a rigorous experience with defined performance standards. In spite of the best efforts of cadre conducting BCT to create a safe environment oriented for success, significant numbers of trainees fail to meet minimal standards or are injured during training. A recent study of BCT at Ft. Jackson, SC demonstrated that among 2,072 trainees who entered training, 19% failed to complete training with their peers,15% failed the Army Physical Fitness Test (APFT) at week 7, and 29% suffered an overuse injury. Women had a higher incidence than men for each of these three negative training outcomes.

Analytic methods commonly employed in the evaluation of medical diagnostic tests to develop clinical prediction rules or test item clusters (TICs) have potential to yield useful combinations of tests for estimating probabilities of negative training outcomes. These methods allow identification of baseline attributes and performance variables that provide discrimination between groups who experience the negative training outcomes and groups who do not experience those outcomes. The identified groupwise predictor variables are entered into a binary logistic regression analysis to find the most parsimonious set of predictors that retains the best possible predictive power for relevant outcomes. This modeling procedure yields the best subset of predictive tests to comprise the TIC. Diagnostic accuracy statistics (sensitivity, specificity, positive likelihood ratios [PLRs], and negative likelihood ratios [NLRs]) are then calculated for both individual predictor variables and for TICs. The likelihood ratios can then be used with individual subjects to reduce the uncertainty about risk depending on the result of the test(s) of prediction. Prediction of an unfavorable outcome does not definitively mean the individual will suffer that outcome because false positives are possible. Likewise, prediction of a favorable outcome does not mean the individual will not suffer a negative outcome because false negatives are possible However, results of individual tests or test clusters can assist in predicting probability of a negative outcome by deriving post-test probability from the pretest probability and the likelihood ratio associated with the test or cluster of tests.

The purpose of this study was to determine whether potentially useful TICs could be derived from an existing data set. We hypothesized that large shifts in pre-test to post-test probability (PLRs >10 or NLRs < 0.1) could result from TICs derived separately for men and women trainees in BCT using a limited set of baseline attribute and performance variables to predict the following undesirable training outcomes:

- failure to pass the APFT at week 7
- any overuse injury during BCT
- failure to complete BCT with peers (for any reason).

This retrospective study used data from a study comparing two physical training regimens evaluated during BCT at Fort Jackson, SC in 2003. In that prior study, a newly developed Standardized Physical Training (SPT) program was evaluated in

comparison to a traditional non-standardized physical training program. We analyzed data from 518 male trainees and 416 female trainees in the SPT group using methods described above. The SPT group was used for analysis because the SPT program was ultimately adopted. Cut scores for continuous predictor variables were determined using receiver-operator characteristic curve analysis with a bias toward minimizing false positives and maximizing PLRs. Among all variables in the data set, 15 potential predictors available at the beginning of BCT were identified for derivation of the TICs. Six TICs were derived to estimate probabilities of APFT failure, overuse injuries, and attrition: each separately for men and women.

APFT FAILURE: WOMEN

Among women in this study with recorded week 7 APFT scores, 58 (18.4%) of 315 trainees with complete data failed the test. For female trainees, 6 baseline predictors (Push-ups Initial Test, Sit-ups Initial Test, Run Time Initial Test, Age, Weight, and Body Mass Index) discriminated between those who passed the APFT at week 7 and those who did not. The logistic regression analysis identified 3 predictors for the TIC: Sit-ups Initial Test (< 10 repetitions), Run Time Initial Test (> 10.71 minutes), and Age (< 20.5 years). The TIC had a PLR as high as 22.77 which would shift an individual female trainee's 18.4% pre-test probability for APFT failure to a post-test probability of 83.7% if all 3 tests in the TIC were positive.

APFT FAILURE: MEN

Among men in this study with recorded week 7 APFT scores, 54 (11.8%) of 459 trainees with complete data failed the test. For male trainees, 8 baseline predictors (Push-ups Initial Test, Sit-ups Initial Test, Run Time Initial Test, Weight, Body Mass Index, Years of Education, Pay Grade and membership in the Fitness Assessment Program) discriminated between those who passed the APFT at week 7 and those who did not. The logistic regression analysis identified 3 predictors for the TIC: Push-ups Initial Test (< 13 repetitions), Sit-ups Initial Test (< 21 repetitions), and membership in the Fitness Assessment Program. The TIC had a PLR as high as 10.47 which would shift an individual male trainee's 11.8% pre-test probability for APFT failure to a post-test probability of 58.4% if 2 of the 3 tests in the TIC were positive.

OVERUSE INJURIES: WOMEN

Among women in this study, 157 (37.7%) of 416 trainees experienced at least one overuse injury. For female trainees, 2 baseline predictors (Push-ups Initial Test and Run Time Initial Test) discriminated between those who experienced one or more overuse injuries and those who did not. The logistic regression analysis identified only 1 predictor for the TIC: Push-ups Initial Test (< 4 repetitions). The Push-ups Initial Test had a PLR of 1.30 which would shift an individual female trainee's 37.7% pre-test probability for overuse injuries to a post-test probability of 44.0%, given a positive result for this test.

OVERUSE INJURIES: MEN

Among men in this study, 81 (15.6%) of 518 trainees experienced at least one overuse injury. For male trainees, 6 baseline predictors (Push-ups Initial Test, Age,

Weight, Body Mass Index, Number of Dependents, and Years of Education) discriminated between those who experienced one or more overuse injuries and those who did not. The logistic regression analysis identified 4 predictors for the TIC: Age (> 25.5 years), BMI (> 31.1 kg/m^2), Number of Dependents (> 2), and Years of Education (< 11.5). The TIC had a PLR as high as 51.96 which would shift an individual male trainee's 15.6% pre-test probability for overuse injuries to a post-test probability of 90.6% if 3 of the 4 tests in the TIC were positive.

ATTRITION: WOMEN

Among women in this study, 121 (29.1%) of 416 trainees were lost to attrition. For female trainees, 5 baseline predictors (Push-ups Initial Test, Sit-ups Initial Test, Run Time Initial Test, Years of Education, and Pay Grade) discriminated between those who completed BCT with their peers and those who did not. The logistic regression analysis identified only 1 predictor for the TIC: Run Time Initial Test (> 13.96 minutes). The Run Time Initial Test had a PLR of 6.16 which would shift an individual female trainee's 29.1% pre-test probability for attrition to a post-test probability of 71.7%, given a positive result for this test.

ATTRITION: MEN

Among men in this study, 76 (14.7%) of 518 trainees were lost to attrition. For male trainees, only 1 baseline predictor (Push-ups Initial Test) discriminated between those who completed BCT with their peers and those who did not. The Push-ups Initial Test had a PLR of 4.84 which would shift an individual male trainee's 14.7% pre-test probability for attrition to a post-test probability of 45.5% given a positive result (< 11 repetitions) for this test.

CONCLUSION

Large and potentially conclusive shifts in pre-test to post-test probability were observed with TICs derived to predict APFT failure for both men and women, and to predict overuse injuries in men. These multivariate models suggest that negative training probabilities as high as 91% might be estimated for individual trainees, given positive results for test item clusters. Moderate probability shifts were seen with the single tests identified to predict BCT attrition for both men and women. No useful model for predicting overuse injuries in women was derived from the methods employed in this study. This study suggests good potential for these analytic methods to derive useful combinations of prognostic tests for predicting negative outcomes in BCT.

Future studies with the goal of TIC development should be planned in sequence. First, prospective studies should collect data on the broadest possible spectrum of known and suspected risk factors for negative training outcomes in order to derive more robust and inclusive TICs. This study contained a limited number of predictive factors; there may be other factors useful for predicting the outcomes. Second, TICs must be validated on a second, independent sample of trainees before recommendation for use. Third, research is needed to study the impact of implementing validated TICs on cost, outcome, and behavior.

INTRODUCTION

Basic Combat Training (BCT) in the U.S. Army is a rigorous experience with defined performance standards. In spite of the best efforts of cadre conducting BCT to create a safe environment oriented for success, significant numbers of trainees fail to meet minimal standards or are injured during training. A recent study of BCT at Ft. Jackson, SC[1] demonstrated that among 2,072 trainees who entered training, 19% failed to complete training with their peers,15% failed the Army Physical Fitness Test (APFT) at week 7, and 29% suffered an overuse injury. Women had a higher incidence than men for each of these three negative training outcomes.

Numerous studies have characterized risk factors for injury in BCT[2-6]. These risk factors are both intrinsic (personal characteristics of the individual) and extrinsic (factors external to the individual). Identified intrinsic risk factors include female gender, high foot arches, knee Q-angle >15^{o}, genu valgus, past ankle sprains, low aerobic fitness, low muscular endurance, high and low extremes of flexibility, low levels of physical activity prior to BCT, cigarette smoking prior to BCT, and older age. Less consistently demonstrated intrinsic risk factors include lower levels of muscular strength, higher body fat or body mass index, and white ethnicity. Multivariate analysis have shown that cigarette smoking prior to BCT, low levels of aerobic fitness and low levels of physical activity prior to BCT are independent injury risk factors. Extrinsic risk factors that have been identified in US Army Basic Combat Training (BCT) include high running mileage, training company, older running shoes, and the summer season. The more running mileage that is performed the higher the likelihood that injuries will occur. There are large differences in injury rates between training companies possibly due to differences in training intensities, especially with regard to physical training. Older running shoes are associated with a higher risk of stress fractures. Seasonal variations in injury rates appear to occur in BCT with higher overall rates in the summer and lower rates in the fall.

A considerable amount of work has been done identifying risk factors for attrition from military basic training but most of these studies are in government technical reports and few appear in the published open literature. One review covers attrition during the first term of service in enlisted Soldier, Sailors and Airmen[7]. For basic training attrition from any service, this review found that demographic and psychosocial risk factors included lower educational attainment, female gender, White ethnicity, lower Armed Forces Qualification Test scores, lower moral character (less conformance to laws, rules and regulations), moral waivers, pre-service job instability, and less time in the Delayed Entry Program. The age-attrition relationship was bimodal with higher attrition when youngest, decreasing in 19-23 year olds, and rising again in older individuals. Attrition for mental health reasons was associated with pre-service physical/sexual abuse, previous mental health counseling, previous treatment with medication, previous psychiatric hospitalization, low motivation, pessimism toward training, depression, lack of self-reliance, and referral to a mental health facility during BCT. Attrition was also higher among those waivered for hearing problems, skin disorders, back disorders, and prior knee injuries. Other health-related risk factors included pre-service injury,

injuries during basic training, a history of prior cigarette smoking, low physical activity prior to service, greater body weight, higher body mass index (BMI) and lower physical fitness. Few of the studies in this review had performed multivariate analysis and the authors recommended a comprehensive study to examine a large number of these factors in a single investigation so their interaction and relative importance could be determined. It was considered possible that a number of these factors interact in ways to decrease (due to multicollinearlity) or increase (due to synergistic effects) their influence on attrition risk.

In contrast to the wide-ranging literature on injuries and attrition from BCT we found no studies that would identify risk for APFT failure. Methods employed in many previous risk factor studies have focused primarily on identifying magnitude of relative risk for individual risk factors rather than development of multivariate models to yield predictive test clusters.

TEST ITEM CLUSTERS

In the health care context, a clinical prediction rule is defined as a clinical tool that quantifies the contributions that various components of the history, physical examination, laboratory results, and imaging studies make toward the diagnosis, prognosis, or likely response to treatment in an individual patient.[8,9] Synonyms for clinical prediction rules include clinical prediction guides, clinical decision rules, and test item clusters (TICs).[10,11] The latter term was selected for use in for this study because of the non-clinical setting for the study.

Analytic methods for developing a TIC were initially employed to assist clinicians with diagnosis of medical diseases and conditions, combining multiple elements of the clinical and laboratory examination.[12-15] However, these methods have been recently used for other purposes: identifying patients who will be most likely to respond favorably to a specific clinical treatment[16-20] stratification of subjects into different risk groups, [21-23] or prognosis for a specific outcome.[24,25]

Development of a TIC involves three steps[9]:

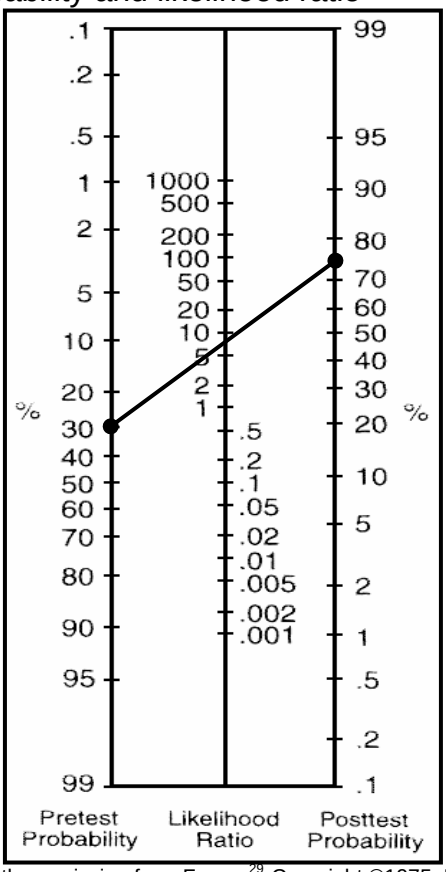
- Step 1: Derivation – Identification of factors with predictive power
- Step 2: Validation – Evidence of reproducible accuracy of the TIC, testing with an independent sample
- Step 3: Impact Analysis – Evidence that implementing the TIC changes behavior and/or improves outcomes

The first step in developing a TIC is to find the most parsimonious set of predictors that retains the best possible predictive power for relevant outcomes. Outcomes of interest must be dichotomous outcomes or outcomes that can be logically dichotomized. Potential predictor variables include all possible variables present at baseline that might logically be able to discriminate between subjects with vs. without outcome(s) of interest. These potential predictor variables can be from categorical or continuous scales.

Once individual potential predictor variables are identified, a TIC is derived using a multivariate modeling technique such as discriminate function analysis, recursive partitioning, or multiple logistic regression.[8] Diagnostic accuracy statistics (sensitivity, specificity, positive likelihood ratios [PLRs], and negative likelihood ratios [NLRs]) are then calculated for both individual predictor variables and for TICs. The likelihood ratios can then be used with individual subjects to reduce the uncertainty about risk depending on the result of the test(s) of prediction.[26,27] Prediction of an unfavorable outcome does not definitively mean the individual will suffer that outcome because false positives are possible. Likewise, prediction of a favorable outcome does not mean the individual will not suffer a negative outcome because false negatives are possible

Results of individual tests or test item clusters can assist in predicting probability of a negative outcome by deriving post-test probability from the pretest probability and the likelihood ratio associated with the test or cluster of tests. Pre-test probabilities are determined in clinical settings based on expertise of the clinician and knowledge of local populations and circumstances. Lacking other information, prevalence of the condition is considered a reasonable estimate of pre-test probability.[10,16] Determining a post-test probability with computational methods requires conversion of the pre-test probability to odds, multiplication of the odds by the likelihood ratio, and conversion of the resulting post-test odds to a post-test probability.[28] A likelihood ratio nomogram proposed initially by Fagan[29] is an alternative graphical method of determining post-test probability. Practical application of these methods can be illustrated with the following example: if the pre-test probability (prevalence) of injury for all female trainees is 30%, and a given female trainee has a positive result for two of three predictive tests in a TIC, then the post-test probability of injury for this individual will shift upward depending on the magnitude of the PLR. If the PLR in this example is 7.5 for two or more positive tests in the TIC, then the probability of injury shifts from 30% to 76.3% for this individual (Figure 1).

Figure 1. Likelihood Ratio Nomogram illustrating determination of post-test probability from pre-test probability and likelihood ratio



Reprinted with permission from Fagan.[29] Copyright ©1975, Massachusetts Medical Society. All rights reserved. Adapted with permission.

The utility of diagnostic tests or predictors can be classified according to the magnitudes of likelihood ratios using the nomenclature proposed by Jaeschke et al.[28] (Table 1):

Table 1. Characterization of diagnostic or predictive tests based on magnitude of likelihood ratios

	PLR[a]	**NLR[b]**
Large and often conclusive[c] shifts from pre- to post-test probability	>10	<0.1
Moderate shifts from pre- to post-test probability	5 - 10	0.1 - 0.2
Small but sometimes important shifts from pre- to post-test probability	2 - 5	0.5 - 0.2
Small and rarely important shifts from pre- to post-test probability	1 - 2	0.5 - 1

[a] PLR = positive likelihood ratio

[b] NLR = negative likelihood ratio

[c] Note that the word “conclusive” here is only a qualitative characterization of the magnitude of the PLRs and NLRs. In any effort of this type further research is likely warranted and additional validation necessary.

OBJECTIVES

The purpose of this study was to determine whether potentially useful TICs could be derived from an existing data set.[1] We hypothesized that large shifts in pre-test to post-test probability (PLRs >10) could result from TICs derived separately for men and women trainees in BCT using a limited set of baseline attribute and performance variables to predict the following undesirable training outcomes: 1) failure to pass the APFT at week 7; 2) any overuse injury during BCT; 3) failure to complete BCT with peers (for any reason).

METHODS

This retrospective study was conducted using data from a study comparing two physical training regimens evaluated during BCT at Fort Jackson, SC in 2003.[1] In that prior study, a newly developed Standardized Physical Training (SPT) program was evaluated in comparison to a traditional non-standardized physical training program. Given that the SPT program was formally adopted and implemented at all Initial Entry Training locations beginning 02 February 2004, we decided to analyze data only from trainees in the SPT group.

SUBJECTS

There were 934 trainees in the SPT group. Of these, 518 (55.5%) were men; 416 (44.5%) were women. Baseline characteristics are summarized in Tables 2 & 3.

Table 2. Baseline attributes for male trainees

Baseline Attribute	Mean or Percentage	Std. Deviation	Minimum	Maximum
Age (years) (years)	21.98	3.94	17	35
Weight (lbs)	173.44	30.22	105	307
Height (in)	69.22	2.82	61	78
BMI (kg/m^2)	25.39	3.97	17.13	37.32
Number of Dependents	0.38	0.82	0	4
Armed Forces Qualification Test	58.80	20.70	28	99
Term of enlistment (years)	4.04	1.07	2	6
Years of Education	12.44	1.32	9	19
Average Household Income ($)	58,125	26,691	25,889	184,834
Push-ups Initial Test (1min)	28.98	11.47	0	63
Sit-ups Initial Test (1min)	30.80	7.13	0	59
Run Time Initial Test (1mile)	8.48	1.51	5.43	20.50
Race: White (%)	61.39	-	-	-
Race: Black (%)	17.76	-	-	-
Race: Hispanic (%)	13.51	-	-	-
Race: Other (%)	7.14	-	-	-
Component: National Guard (%)	13.51	-	-	-
Component: Regular Army (%)	72.01	-	-	-
Component: Army Reserve (%)	14.48	-	-	-
Education: Less than High School (%)	0.19	-	-	-
Education: High School Graduate (%)	48.26	-	-	-
Education: GED (%)	7.53	-	-	-
Education: Some College (%)	8.88	-	-	-
Education: College Graduate (%)	4.83	-	-	-
Education: Unknown (%)	30.31	-	-	-

Table 3. Baseline attributes for female trainees

Baseline Attribute	Mean or Percentage	Std. Deviation	Minimum	Maximum
Age (years) (years)	21.81	4.18	17	36
Weight (lbs)	139.26	21.07	93	202
Height (in)	64.43	2.56	59	72
BMI (kg/m^2)	23.51	2.91	16.67	32.24
Number of Dependents	0.41	0.85	0	5
Armed Forces Qualification Test	54.42	18.45	31	99
Term of enlistment (years)	3.99	0.90	2	6
Years of Education	12.35	1.11	11	16
Average Household Income ($)	53,163	17,651	25,889	135,315
Push-ups Initial Test (1min)	9.38	8.72	0	60
Sit-ups Initial Test (1min)	24.11	8.91	0	46
Run Time Initial Test (1mile)	10.52	1.62	6.23	17.92
Race: White (%)	49.76	-	-	-
Race: Black (%)	26.68	-	-	-
Race: Hispanic (%)	14.42	-	-	-
Race: Other (%)	9.13	-	-	-
Component: National Guard (%)	15.38	-	-	-
Component: Regular Army (%)	74.52	-	-	-
Component: Army Reserve (%)	10.10	-	-	-
Education: Less than High School (%)	0.48	-	-	-
Education: High School Graduate (%)	54.57	-	-	-
Education: GED (%)	6.73	-	-	-
Education: Some College (%)	8.65	-	-	-
Education: College Graduate (%)	3.61	-	-	-
Education: Unknown (%)	25.96	-	-	-

BASIC COMBAT TRAINING

Phases of Training

Basic Training was divided into three phases, each about 3 weeks in duration. All phases included non tactical road marches of varying length to and from training sites in which trainees marched in formation. Red Phase (Patriot Phase) consisted of introductory lessons in customs and courtesies, drill and ceremony, physical fitness, nutrition, first aid, wearing of the uniform, rifle maintenance, the manual of arms, and radio/telephone communication procedures. Red phase was characterized by total cadre control and constant supervision. Army values were introduced and reinforced throughout training. Major physical training events included Victory Tower, introductory tactical road march, introduction to bayonet training, and the conditioning obstacle course.

The second phase (Gunfighter or White Phase) placed emphasis on basic rifle marksmanship (BRM); 14 BRM lessons and a final test were required to qualify with the M16 rifle. Training on the M60 machine gun, M203 grenade launcher and M18 Claymore mine were also provided. Major events requiring physical activity included continued bayonet training (including pugil training), nuclear, biological, and chemical (NBC) defense, hand to hand combat, two tactical foot marches, and continued drill and ceremony training.

The final phase (Warrior or Blue Phase) was designed to teach individual tactical skills and emphasize the importance of teamwork. Combat maneuver, live fire exercises, and a three day field training exercise (FTX, Victory Forge) were conducted. Major physical activities included the hand grenade qualification range, individual tactical training, the confidence course, conditioning obstacle course, and the FTX. For the FTX, trainees spent three days in the field demonstrating proficiency in common military skills. Soldiers participated in a graduation ceremony before moving on to their advanced individual training (AIT) sites.

A typical training day began at 0530 with a wake-up by the drill sergeant. The trainees dressed in PT uniform and performed PT for 1-1.5 hours. After PT, trainees returned to the barracks, changed into BDUs, had a formation, and filed into the mess hall for breakfast. After breakfast, the training events of the day were conducted. Often these involved non-tactical road marches or motorized transportation to field training sites or classroom instruction in the battalion area. Lunch was generally served at 1200, either in the battalion mess hall or in the field. Training continued in the afternoon with dinner at about 1700. Generally training continued until about 2030, Trainees had personal time from 2030 to 2130 when lights went out. Generally, no training was conducted on Sunday.

Physical Fitness Tests

During BCT trainees took 3 scheduled fitness tests. The first test was the Fitness Assessment taken within 1-3 days of arrival. The Fitness Assessment consisted of a 1-minute maximal effort push-ups event, a 1-minute maximal effort sit-up event, and a 1-mile run for time. This was also called the 1/1/1Test. The other 2 tests involved the standard Army Physical Fitness Test (APFT) taken at Weeks 5 and 7. The APFT consisted of a 2-minute maximal effort push-ups event, a 2-minute maximal effort sit-up event, and a 2-mile run for time. This was also called the 2/2/2 Test. Both fitness tests were administered by the drill sergeants who were very familiar with the well-standardized test procedures.

The final APFT given on Week 7 (see Figure 2) was the one trainees had to "pass" to meet a mandated BCT graduation requirement. To "pass" the APFT, all trainees were required to meet certain age and gender adjusted criteria involving obtaining a minimum of 50 age- and gender-adjusted "points" on each test event.[30] A trainee who obtained 100 points on 2 events but 49 points on the third event was considered an APFT failure.

AVAILABLE VARIABLES

The analysis was limited to the set of available variables in the existing data set. Three outcome variables and 15 potential predictor variables were identified.

Outcome Variables

APFT Failure. Passing the final APFT in week 7 of BCT was the standard expectation for all trainees. Failure of the week 7 APFT was defined as a negative outcome of interest for this study, even though some trainees who failed the week 7 APFT subsequently passed APFT retakes offered during weeks 7-9. It is considered very important for trainees to "pass" the test at Week 7 because training is designed for this outcome and because administration of additional tests is very time consuming and interferes with other training activities.

Overuse Injuries. Each outpatient encounter with a trainee at the Fort Jackson Troop Medical Center (TMC) or Hospital resulted in data entry into the Standard Ambulatory Data Record (SADR), including diagnostic codes from the International Classification of Diseases, Version 9 (ICD-9). Data from the SADR including ICD-9 codes were uploaded into a database maintained by the Army Medical Surveillance Activity (AMSA). From the AMSA database, IDC-9 codes for all medical visits of all trainees in the study were extracted. Trainees having one or more specific ICD-9 codes comprising the Overuse Injury Index (OII) were identified. The OII attempts to capture musculoskeletal injuries resulting from cumulative microtrauma (overuse type injuries). A full listing of ICD-9 codes comprising the OII is provided in Appendix I of the SPT evaluation study report.[1] One or more TMC or hospital visits for any injury included in the OII was the operational definition for presence of an overuse injury during BCT for the current study. This definition underrepresented the actual incidence of overuse injuries because visits to Battalion Aid Stations were not entered into the SADR.

Attrition. Full cycle trainees were those that began training the first day of the company training cycle and graduated from BCT with that same company 9 weeks later. Trainees who began but did not finish with their units were considered to have attrited. Trainees could be lost to attrition in two major ways: discharge or newstart. A discharged trainee was one who was not suitable for service in the Army and was formally released from his or her service commitment. There were numerous reasons a trainee could be discharged but most reasons fell into two major categories: medical conditions that existed prior to service (EPTS discharge) or poor entry-level performance. The latter category was often called an entry-level separation (ELS) or Chapter 11 discharge. ELS discharges were most often the result of the trainee's inability to adapt to the military environment because of lack of ability (cannot adequately perform critical military tasks) or for psychosocial reasons (motivation, inability to follow orders, personality problems, etc.). A newstart was a trainee who was "recycled" for inability to complete mandatory requirements with peers for reasons such as motivation, injury, emergency leave, or inability to meet specific training standards (i.e., difficulty developing specific skills like basic rifle marksmanship). Newstarts were sent to another unit to be given one or more additional opportunities to fulfill BCT requirements for graduation. Once trainees left the unit under study they were considered to have attrited for the purposes of this investigation.

Potential Predictor Variables

Continuous-Scale Variables

- Push-ups Initial Test. The Push-ups Initial Test was the first element of the Fitness Assessment given to all trainees during the first week of BCT. For the Push-ups Initial Test, a trainee was required to lower the body from the front-leaning rest position in a generally straight line to a point where the upper arms were parallel to the ground, and then return to the starting position with the elbows fully extended. For this test, trainees wore the improved physical fitness uniform with socks and running shoes. The maximum number of correct repetitions a trainee could perform in 1 minute was recorded and served as the Push-up Initial Test score.

- Sit-ups Initial Test. The Sit-ups Initial Test was the second element of the Fitness Assessment given to all trainees during the first week of BCT. For this test, the trainee lay supine on the ground, knees flexed to a 90° angle, and fingers were interlocked behind the head. A second person held the trainee's ankles and kept the trainee's heels firmly on the ground. The trainee raised the upper body to a vertical position so that the base of the neck was above the base of the spine; the trainee then returned to the starting position. For this test, trainees wore the improved physical fitness uniform with socks and running shoes. The maximum number of correct repetitions a trainee could perform in 1 minute was recorded and served as the Sit-Up Initial Test score.

- Run Time Initial Test. The Run Time Initial Test was the third element of the Fitness Assessment given to all trainees during the first week of BCT. Trainees ran or walked as fast as possible over a 1-mile course while wearing the improved physical fitness uniform with socks and running shoes. Time to complete the 1-mile course was recorded and served as the Run Time Initial Test score.

- Age. Age in years was obtained for each trainee from a database management system called the Warrior Training Room (WTR). WTR was maintained by the Training Non-Commissioned Officer (NCO) of each company.

- Weight. Weight in pounds was obtained for each trainee from a database management system called the Reception Battalion Automated Support System (RECBASS).

- Height. Height in inches was obtained for each trainee from the RECBASS.

- Body Mass Index. Body Mass Index (BMI) was calculated from height and weight using the formula: weight/height^2 (kg/m^2).

- Number of Dependents. Number of Dependents was obtained for each trainee from the US Army Accessions Command Regular Army Data Warehouse.

- Armed Forces Qualification Test. The Armed Forces Qualification Test (AFQT) Score was obtained for each trainee from the US Army Accessions Command Regular Army Data Warehouse. The AFQT was designed to measure the trainability of recruits. The AFQT score was derived as the sum of the standardized scores from several elements of the Armed Service Vocational Aptitude Battery: Arithmetic Reasoning plus Math Knowledge plus twice the sum of the Paragraph Comprehension and Word Knowledge subtests. Scores were percentiles ranging from 10-99 (the lowest 10 percentiles were excluded from military service).[31]

- Years of Education. Years of Education was obtained for each trainee from the US Army Accessions Command Regular Army Data Warehouse.

- Average Household Income. Average Household Income was obtained for each trainee from the US Army Accessions Command Regular Army Data Warehouse.

Categorical-Scale Variables

- Pay Grade. Pay Grade (E-1, E-2, E-3, or E-4) was obtained for each trainee from the WTR.

- Race. Race was obtained for each trainee from the WTR. Race was self-reported as one of the following categories: Asian, Black, Caucasian, Hispanic, Native American, Other.

- Component. Component of military service was obtained for each trainee from WTR. Component was recorded as one of the following categories: National Guard, Regular Army, Army Reserve.

- Fitness Assessment Program. Historically, when a new trainee arrived for BCT they took the Reception Station Physical Fitness Test (RSPFT). If they failed this test they entered the Fitness Assessment Program (FAP) where they physically trained under the guidance of drill sergeants until they could pass the test. The criteria for passing the test are shown in Table 4. For this study, trainees who failed the RSPFT entered BCT *without* participating in the FAP. Thus, FAP status indicates very low physical fitness on entry. Status regarding

trainees who should have entered the FAP was obtained for each trainee from a database contained in the FAP company orderly room.

Table 4. Passing standards for the Reception Station Physical Fitness Test

Event	Men	Women
Push-ups (repetitions)	13	3
Sit-ups (repetitions)	17	17
1-mile Run Time (minutes)	8.5	10.5

DATA ANALYSIS

Eight TICs were derived to estimate probabilities of 1) APFT failure at Week 7, 2) overuse injuries, and 3) attrition. The TICS were derived separately for men and women for each of the 3 outcomes, with 2 additional TICs for overuse injuries using Cox regression methods. Deriving the predictive models involved multiple steps described below.

Groupwise Discrimination for Individual Predictors

First level analyses of between-group differences served as a crude filter to determine whether potential predictor variables could discriminate between groups of trainees with the outcomes of interest vs. those without. These analyses were performed with unpaired t-tests for continuous predictor variables and Chi-square tests for categorical variables. If the sample size is small to moderate, it is common to relax alpha to 0.10 or 0.15 to avoid Type II errors at this stage of the process. However, because the size of this sample was more than adequate given the number of potential predictor variables,[32] we used an alpha level of 0.05 to protect against Type I error.

Prognostic Accuracy for Individual Predictors

Sensitivity, specificity, positive and negative likelihood ratios were calculated for each predictor variable that discriminated ($p \leq 0.05$) between groups of subjects with vs. without negative training outcomes. For any variable yielding a frequency count of zero for a cell of the 2 x 2 table for computation of prognostic accuracy statistics, a value of 0.5 was added to all four cells.[33] This process was straightforward for categorical variables that were already dichotomized, allowing immediate preparation of 2 x 2 tables for cross-tabulation. For continuous variables and for categorical variables with more than two levels, intermediate steps were required to create dichotomous scores.

Dichotomizing Categorical-Scale Predictors. Categorical predictor variables with more than two variables were dichotomized so that the 2 x 2 tables

could be constructed with frequency counts for computation of prognostic accuracy statistics. This was accomplished ad hoc by identifying the one category that had greatest intuitive appeal as the category of highest risk, and by collapsing all remaining categories into a single "other" category. Attention was given to the resulting frequency counts to avoid creating empty cells or cells with very small counts.

Dichotomizing Continuous-Scale Predictors. Continuous predictor variables were dichotomized by establishing a cut score with receiver-operator curve (ROC curve) analysis. This process computes sensitivity and specificity for multiple cut scores along the continuum of the scale, yielding coordinates for a plot so that the characteristics of the scale can be observed graphically. The plot was constructed with sensitivity on the y-axis and 1-specificity on the x-axis. This allowed visualization of the ability of the scale to maximize true positives while minimizing false positives across the spectrum of possible cut scores for the scale. An ideal diagnostic or predictive test would have a very steep initial slope, an abrupt transition point, and then a less steep slope beyond that point. Area under the ROC curve (AUC) was used as one measure of how "ideal" each continuous-scale predictor variable was in this respect . Because no variable approached the ideal, a decision was made to minimize the proportion of false positives by selecting a cut score with high specificity and high positive likelihood ratio.

Once all predictor variables were dichotomized, 2 x 2 tables were constructed containing frequency counts expressing numbers of trainees with true positive test results, false positive test results, true negative test results, and false negative test results. For each individual predictor, we calculated the post-test probability of the outcome based on the pre-test probability (prevalence of the outcome in the sample), and the likelihood ratio.[34]

Multivariate Predictive Model Derivation

Binary logistic regression analysis was used to filter the set of predictor variables further and to derive a multivariate model (TIC) that eliminated redundant or substantially correlated predictors, or any predictors that did not contribute meaningfully to the multivariate prediction. The goal of this process was to yield a parsimonious set of predictor variables that would be logical and consistent and that could provide independent information about the likelihood of an outcome. Potential predictors that yielded p-values < 0.05 from the t-tests and Chi-square tests were entered into the logistic regression analysis using a forward stepwise procedure. The predictor variables chosen for retention by the forward stepwise method all had significant changes ($p \leq 0.05$) in -2 log-likelihood of the model when added from the previous step of model development. For continuous-scale variables and categorical-scale variables with more than two levels, the raw (non-dichotomized) values were entered into the logistic regression analysis.

Cox regression procedures were used to construct potential alternative TICs for overuse injuries. We wanted to explore whether TICs based on Cox regression

methods might peform better than TICs based on logistic regression methods for the outcome associated with unequal time at risk in trainees that attrited. For Cox regression modeling, survival days until first overuse injury was specified as the time variable; the status variable was presence or absence of overuse injury. Potential predictors that yielded p-values less than 0.05 from the t-tests and Chi-square tests were entered as model covariates. The predictor variables chosen for retention by the Cox regression method all generated significant changes ($p \leq 0.05$) in -2 log-likelihood of the model when added from the previous step of model development. For continuous-scale variables and categorical-scale variables with more than two levels, the raw (non-dichotomized) values were entered into the Cox regression analysis.

Predictors retained by the logistic regression and Cox regression models comprised the TICs. Each TIC was characterized further with calculation of sensitivity, specificity, positive and negative likelihood ratios for each level of positive predictors in the cluster. For example, if the model yielded a set of 3 predictors, we calculated prognostic accuracy statistics for 3 levels of the TIC: any 1 or more positive predictors, any 2 or more positive predictors, and all 3 positive predictors. For each level of positive predictors in the TIC, we calculated the post-test probability of the outcome based on the pre-test probability (prevalence) and the likelihood ratio.

RESULTS

APFT FAILURE: WOMEN

Among women in this study with recorded week 7 APFT scores, 58 (18.4%) of 315 trainees with complete data failed the test. For female trainees, 6 continuous-scale predictors (Table 5) and no categorical-scale predictors (Table 6) discriminated ($p \leq 0.05$) between those who passed the APFT at week 7 and those who did not. These predictors were Push-ups Initial Test, Sit-up Initial Test, Run Time Initial Test, Age, Weight and BMI. Weight was not entered into the logistic regression analysis because weight is a determinant of BMI.

Table 5. Comparisons of means: APFT success vs. failure for female trainees

Potential Predictor	APFT performance	N	Mean	Std. Deviation	t	df	p-value (2-tailed)
Push-ups Initial Test*	pass	256	11.23	9.07	6.17†	112.28†	<0.001†
	fail	56	5.00	6.25			
Sit-ups Initial Test*	pass	256	26.03	8.49	6.73	310.00	<0.001
	fail	56	17.63	8.36			
Run Time Initial Test*	pass	255	10.13	1.40	-4.59	309.00	<0.001
	fail	56	11.11	1.68			
Age (years)*	pass	257	22.07	4.23	3.25†	112.84†	0.001†
	fail	58	20.52	3.04			
Weight (lbs)	pass	257	137.43	20.03	-2.95	313.00	0.003
	fail	58	146.19	22.09			
Height (in)	pass	257	64.28	2.43	-0.59†	74.65†	0.556†
	fail	58	64.53	3.01			
BMI (kg/m^2)*	pass	257	23.32	2.84	-3.12	313.00	0.002
	fail	58	24.61	2.89			
Number of Dependents	pass	185	0.44	0.86	1.67†	88.40†	0.098†
	fail	44	0.25	0.61			
Armed Forces Qualification Test	pass	185	56.10	19.15	1.55	227.00	0.124
	fail	44	51.20	17.80			
Years of Education	pass	185	12.45	1.20	1.21†	82.00†	0.230†
	fail	44	12.25	0.92			
Average Household Income ($)	pass	181	53895.54	19373.22	0.16	221.00	0.877
	fail	42	53395.90	15866.21			

*Selected for entry into logistic regression analysis

†Adjusted for heterogeneity of variance

Table 6. Comparisons of frequencies: APFT success vs. failure for female trainees (Chi-Square test results)

	Pearson Chi-Square	df	p-value (2-sided)
Pay Grade	6.63	3	0.085
Race	6.78	5	0.237
Component	1.12	2	0.573
Fitness Assessment Program	1.93	1	0.165

Table 7 contains derived cut score values, prognostic accuracy statistics, associated pre-test to post-test probability shifts, and results of ROC curve analyses for the 5 predictor variables selected for entry into the logistic regression model. Plots of ROC curves are displayed in Appendix A. An example of coordinate points for an ROC curve illustrating selection of a cut score maximizing the positive likelihood ratio is presented in Appendix B.

Table 7. Performance of selected single variables in prediction of APFT failure for female trainees

	Cut Score: positive test	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability	AUC	p
Push-ups Initial Test	<2	0.48	0.85	3.25	0.61	18.4%	42.3%	12.1%	0.724	<0.001
Sit-ups Initial Test*	<10	0.23	0.96	6.60	0.80	18.4%	59.8%	15.2%	0.767	<0.001
Run Time Initial Test*	>10.71	0.57	0.71	1.97	0.60	18.4%	30.7%	12.0%	0.685	<0.001
Age (years)*	<20.5	0.69	0.52	1.44	0.60	18.4%	24.5%	11.8%	0.606	0.012
BMI (kg/m^2)	>24.76	0.53	0.70	1.76	0.67	18.4%	28.4%	13.1%	0.629	0.002

*Selected for retention in TIC by logistic regression analysis

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio; AUC = area under the ROC curve; p = p-value for the null hypothesis test that the true AUC = 0.50.

The logistic regression model retained 3 of the 5 entered predictors: Sit-ups Initial Test, Run Time Initial Test, and Age. The model was statistically significant ($p < 0.001$) and yielded a Nagelkerke R^2 value of 0.314. Performance of the TIC with the 3 predictors retained by the logistic regression model is presented in Table 8. Cross-tabulations for each level of the TIC are presented in Tables 9-11. Prior to providing data for the TIC, any female trainee had a pre-test probability of 18.4% for APFT failure based on the prevalence of APFT failure among all female trainees. A female trainee with positive scores for all 3 predictors in the TIC had a post-test probability of 83.7% for APFT failure.

Table 8. Performance of levels in the TIC for prediction of APFT failure for female trainees

	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability
Any 1 or more tests positive	0.91	0.36	1.42	0.25	18.4%	24.3%	5.3%
Any 2 or more tests positive	0.52	0.84	3.14	0.58	18.4%	41.5%	11.5%
All 3 tests positive	0.09	1.00	22.77	0.91	18.4%	83.7%	17.1%

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio.

Table 9. Frequencies of female trainees who failed vs. passed the APFT with any 1 or more tests positive vs. no tests positive in the TIC

	Failed APFT	Passed APFT
Any 1 or more tests positive	51	163
No tests positive	5	92

Percent correct classification: 46.0%

Table 10. Frequencies of female trainees who failed vs. passed the APFT with any 2 or more tests positive vs. less than 2 tests positive in the TIC

	Failed APFT	Passed APFT
Any 2 or more tests positive	29	42
Less than 2 tests positive	27	213

Percent correct classification: 77.8%

Table 11. Frequencies of female trainees who failed vs. passed the APFT with all 3 tests positive vs. less than 3 tests positive in the TIC

	Failed APFT	Passed APFT
All 3 tests positive	5	1
Less than 3 tests positive	51	254

Percent correct classification: 83.3%

APFT FAILURE: MEN

Among men in this study with recorded week 7 APFT scores, 54 (11.8%) of 459 trainees with complete data failed the test. For male trainees, 6 continuous-scale predictors (Table 12) and 2 categorical-scale predictors (Table 13) discriminated ($p \leq 0.05$) between those who passed the APFT at week 7 and those who did not. These predictors were Push-ups Initial Test, Sit-up Initial Test, Run Time Initial Test, Weight, BMI, Years of Education, Pay Grade, and Fitness Assessment Program. Here again, weight was not entered into the logistic regression analysis because weight is a determinant of BMI. The variable Years of Education was also not entered into the regression analysis because the ROC AUC was not significantly greater than 0.5.

Table 12. Comparisons of means: APFT success vs. failure for male trainees

Potential Predictor	APFT performance	N	Mean	Std. Deviation	t	df	p-value (2-tailed)
Push-ups Initial Test*	pass	404	30.79	10.67	8.00	456.00	<0.001
	fail	54	18.44	10.47			
Sit-ups Initial Test*	pass	404	31.62	6.81	5.70	456.00	<0.001
	fail	54	26.06	6.27			
Run Time Initial Test*	pass	404	8.37	1.49	-4.43	455.00	<0.001
	fail	53	9.32	1.43			
Age (years)	pass	405	22.07	3.97	1.07	457.00	0.287
	fail	54	21.46	3.58			
Weight (lbs)	pass	405	171.79	28.60	-3.66	457.00	<0.001
	fail	54	187.44	35.86			
Height (in)	pass	405	69.08	2.81	-1.94	457.00	0.052
	fail	54	69.87	2.66			
BMI (kg/m^2)*	pass	405	25.25	3.80	-2.52‡	62.47‡	0.014‡
	fail	54	26.94	4.73			
Number of Dependents	pass	282	0.38	0.82	0.43	318.00	0.671
	fail	38	0.32	0.77			
Armed Forces Qualification Test	pass	282	60.23	20.80	1.92	318.00	0.056
	fail	38	53.34	20.77			
Years of Education	pass	282	12.50	1.41	2.75‡	81.37‡	0.007‡
	fail	38	12.11	0.73			
Average Household Income ($)	pass	273	58751.65	28125.75	1.35	308.00	0.179
	fail	42	52333.59	18487.31			

*Selected for entry into logistic regression analysis

‡Adjusted for heterogeneity of variance

Table 13. Comparisons of frequencies: APFT success vs. failure for male trainees (Chi-Square test results)

	Pearson Chi-Square	df	p-value (2-sided)
Pay Grade*	7.87	3	0.049
Race	10.73	6	0.097
Component	1.24	2	0.538
Fitness Assessment Program*	18.15	1	<0.001

*Selected for entry into logistic regression analysis

Table 14 contains derived cut score values, prognostic accuracy statistics, associated pre-test to post-test probability shifts, and results of ROC curve analyses for the 6 predictor variables selected for entry into the logistic regression model.

Table 14. Performance of selected single variables in prediction of APFT failure for male trainees

	Cut Score: positive test	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability	AUC	p
Push-ups Initial Test*	<13	0.31	0.97	10.60	0.71	11.8%	58.6%	8.6%	0.795	<0.001
Sit-ups Initial Test*	<21	0.20	0.95	4.11	0.84	11.8%	35.5%	10.1%	0.734	<0.001
Run Time Initial Test	>9.39	0.53	0.84	3.23	0.56	11.8%	30.2%	7.0%	0.704	<0.001
BMI (kg/m^2)	>27.71	0.48	0.74	1.82	0.70	11.8%	19.6%	8.6%	0.612	0.008
Pay Grade	<E-2	0.67	0.49	1.30	0.68	11.8%	14.9%	8.4%	NA	NA
In FAP Program*	yes	0.24	0.93	3.61	0.81	11.8%	32.6%	9.8%	NA	NA

*Selected for retention in TIC by logistic regression analysis

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio; AUC = area under the ROC curve; p = p-value for the null hypothesis test that the true AUC = 0.50; NA = not applicable (categorical variables)

The logistic regression model retained 3 of the 6 entered predictors: Push-ups Initial Test, Sit-ups Initial Test, and membership in the FAP. The model was statistically significant ($p < 0.001$) and yielded a Nagelkerke R^2 value of 0.283. Performance of the TIC with the 3 predictors retained by the logistic regression model is presented in Table 15. Calculations for TIC performance for all 3 tests positive were not completed because there were no male trainees with all 3 tests positive in the sample. Cross-tabulations for in the first two levels of the TIC are presented in Tables 16-17. Prior to providing data for the TIC, any male trainee had a pre-test probability of 11.8% for APFT failure based on the prevalence of APFT failure among all male trainees. A male trainee with positive scores for any 2 of the 3 predictors in the TIC had a post-test probability of 58.4% for APFT failure.

Table 15. Performance of levels in the TIC for prediction of APFT failure for male trainees

	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability
Any 1 or more tests positive	0.63	0.87	4.80	0.43	11.8%	39.1%	5.4%
Any 2 tests positive	0.13	0.99	10.47	0.88	11.8%	58.4%	10.5%
All 3 tests positive	*	*	*	*	*	*	*

*Not computed because no male trainees had all 3 tests positive

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio.

Table 16. Frequencies of male trainees who failed vs. passed the APFT with any 1 or more tests positive vs. no tests positive in the TIC

	Failed APFT	Passed APFT
Any 1 or more tests positive	34	53
No tests positive	20	351

Percent correct classification: 84.1%

Table 17. Frequencies of male trainees who failed vs. passed the APFT with any 2 tests positive vs. less than 2 tests positive in the TIC

	Failed APFT	Passed APFT
Any 2 tests positive	7	5
Less than 2 tests positive	47	399

Percent correct classification: 88.6%

OVERUSE INJURIES: WOMEN

Among women in this study, 157 (37.7%) of 416 trainees experienced at least one overuse injury. For female trainees, 2 continuous-scale predictors (Table 18) and no categorical-scale predictors (Table 19) discriminated ($p \leq 0.05$) between those who experienced one or more overuse injuries and those who did not. These predictors were Push-ups Initial Test and Run Time Initial Test.

Table 18. Comparisons of means: presence vs. absence of overuse injuries for female trainees

Potential Predictor	Overuse Injuries	N	Mean	Std. Deviation	t	df	p-value (2-tailed)
Push-ups Initial Test*	no	254	10.13	9.40	2.26†	373.04†	0.017†
	yes	150	8.11	7.30			
Sit-ups Initial Test	no	254	24.50	8.81	1.17	402	0.243
	yes	150	23.43	9.05			
Run Time Initial Test*	no	251	10.40	1.66	-1.96	398	0.051
	yes	149	10.73	1.55			
Age (years)	no	259	21.62	4.14	-1.21	414	0.226
	yes	157	22.13	4.23			
Weight (lbs)	no	259	139.72	20.78	0.57	414	0.566
	yes	157	138.50	21.59			
Height (in)	no	259	64.37	2.49	-0.60	414	0.548
	yes	157	64.52	2.66			
BMI (kg/m^2)	no	259	23.63	2.85	1.09	414	0.275
	yes	157	23.31	3.00			
Number of Dependents	no	191	0.37	0.84	-1.07	306	0.286
	yes	117	0.48	0.88			
Armed Forces Qualification Test	no	191	55.14	19.20	0.87	306	0.386
	yes	117	53.26	17.17			
Years of Education	no	191	12.37	1.17	0.36	306	0.720
	yes	117	12.32	1.02			
Average Household Income ($)	no	187	54501.25	17239.68	1.69	298	0.091
	yes	113	50948.17	18173.22			

*Selected for entry into logistic regression analysis and Cox regression analysis

†Adjusted for heterogeneity of variance

Table 19. Comparisons of frequencies: presence vs. absence of overuse injuries for female trainees (Chi-Square test results)

	Pearson Chi-Square	df	p-value (2-sided)
Pay Grade	6.97	3	0.073
Race	4.39	5	0.495
Component	0.90	2	0.638
Fitness Assessment Program	1.75	2	0.417

Table 20 contains derived cut score values, prognostic accuracy statistics, associated pre-test to post-test probability shifts, and results of ROC curve analyses for the 2 predictor variables selected for entry into the two regression models.

Table 20. Performance of selected single variables in prediction of overuse injuries for female trainees

	Cut Score: positive test	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability	AUC	p
Push-ups Initial Test*	<4	0.37	0.71	1.30	0.88	37.7%	44.0%	34.7%	0.553	0.076
Run Time Initial Test‡	>10.54	0.52	0.65	1.49	0.73	37.7%	47.5%	30.7%	0.588	0.043

*Selected for retention in TIC by logistic regression analysis

‡Selected for retention in TIC by Cox regression analysis

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio; AUC = area under the ROC curve; p = p-value for the null hypothesis test that the true AUC = 0.50.

The logistic regression model retained only 1 of the 2 entered predictors: Push-ups Initial Test. The model was statistically significant (p = 0.026) and yielded a Nagelkerke R^2 value of 0.017. Performance of the Push-ups Initial Test as a univariate predictor of overuse injuries in women is presented above in Table 20. A cross-tabulation of frequencies for the Push-ups Initial Test is presented in Table 21. Prior to providing data for prediction, any female trainee had a pre-test probability of 37.7% for overuse injuries based on the prevalence of overuse injuries among all female trainees. A female trainee who performed less than 4 correct repetitions on the Push-ups Initial Test had a post-test probability of 44.0% for overuse injuries.

Table 21. Frequencies of female trainees who experienced vs. did not experience overuse injuries with a positive test (<4 push-ups) vs. a negative test (≥ 4 push-ups) for the Push-ups Initial Test

	Overuse Injury	No Overuse Injury
<4 push-ups	56	73
≥ 4 push-ups	94	181

Percent correct classification: 58.7%

The Cox regression model retained only 1 of the 2 entered predictors: Run Time Initial Test. The model was statistically significant ($p = 0.012$). Performance of the Run Time Initial Test as a univariate predictor of overuse injuries in women is presented above in Table 20. A cross-tabulation of frequencies for the Run Time Initial Test is presented in Table 22. Prior to providing data for prediction, any female trainee had a pre-test probability of 37.7% for overuse injuries based on the prevalence of overuse injuries among all female trainees. A female trainee who could not complete the Run Time Initial Test in less than or equal to 10.54 minutes had a post-test probability of 47.5% for overuse injuries.

Table 22. Frequencies of female trainees who experienced vs. did not experience overuse injuries with a positive test (>10.54 minutes) vs. a negative test (≤ 10.54 minutes) for the Run Time Initial Test

	Overuse Injury	No Overuse Injury
Initial Run >10.54 min.	78	88
Initial Run ≤10.54 min.	71	163

Percent correct classification: 60.3%

OVERUSE INJURIES: MEN

Among men in this study, 81 (15.6%) of 518 trainees experienced at least one overuse injury. For male trainees, 6 continuous-scale predictors (Table 23) and no categorical-scale predictors (Table 24) discriminated ($p \leq 0.05$) between those who experienced one or more overuse injuries and those who did not. These predictors were Push-ups Initial Test, Age, Weight, BMI, Number of Dependents, and Years of Education. Here again, weight was not entered into the logistic regression analysis because weight is a determinant of BMI.

Table 23. Comparisons of means: presence vs. absence of overuse injuries for male trainees

Potential Predictor	Overuse Injuries	N	Mean	Std. Deviation	t	df	p-value (2-tailed)
Push-ups Initial Test*	no	434	29.45	11.71	2.44†	126.50†	0.016†
	yes	81	26.46	9.81			
Sit-ups Initial Test	no	434	30.93	7.20	0.94	513.00	0.345
	yes	81	30.11	6.71			
Run Time Initial Test	no	434	8.45	1.53	-1.04	512.00	0.300
	yes	80	8.64	1.36			
Age (years)*	no	437	21.78	3.75	-2.25†	99.80†	0.027†
	yes	81	23.02	4.70			
Weight (lbs)	no	437	171.98	28.45	-2.12†	97.64†	0.036†
	yes	81	181.32	37.63			
Height (in)	no	437	69.19	2.81	-0.58	516.00	0.564
	yes	81	69.38	2.91			
BMI (kg/m^2)*	no	437	25.23	3.89	-2.21	516.00	0.028
	yes	81	26.28	4.25			
Number of Dependents*	no	309	0.32	0.73	-2.60†	58.13†	0.012†
	yes	52	0.75	1.15			
Armed Forces Qualification Test	no	309	59.47	21.12	1.70†	77.90†	0.092†
	yes	52	54.83	17.62			
Years of Education*	no	309	12.50	1.37	2.86†	95.42†	0.005†
	fail	38	12.11	0.73			
Average Household Income ($)	pass	302	57582.13	25530.24	-.095	349.00	0.345
	fail	49	61469.55	33077.68			

*Selected for entry into logistic regression analysis and Cox regression analysis
†Adjusted for heterogeneity of variance

Table 24. Comparisons of frequencies: presence vs. absence of overuse injuries for male trainees (Chi-Square test results)

	Pearson Chi-Square	df	p-value (2-sided)
Pay Grade	3.85	3	0.279
Race	8.83	6	0.183
Component	3.39	2	0.184
Fitness Assessment Program	2.56	2	0.279

Table 25 contains derived cut score values, prognostic accuracy statistics, associated pre-test to post-test probability shifts, and results of ROC curve analyses for the 5 predictor variables selected for entry into the two regression models.

Table 25. Performance of selected single variables in prediction of overuse injuries for male trainees

	Cut Score: positive test	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability	AUC	p
Push-ups Initial Test	<22	0.35	0.76	1.44	0.86	15.6%	21.1%	13.7%	0.583	0.017
Age (years)*‡	>25.50	0.30	0.86	2.19	0.81	15.6%	28.9%	13.1%	0.569	0.050
BMI (kg/m^2)*	>31.05	0.15	0.93	2.02	0.92	15.6%	27.2%	14.5%	0.574	0.034
Number of Dependents*‡	>2	0.15	0.97	5.94	0.87	15.6%	52.3%	13.8%	0.587	0.044
Years of Education*‡	<11.50	0.15	0.93	2.16	0.91	15.6%	28.5%	14.4%	0.588	0.043

*Selected for retention in TIC by logistic regression analysis

‡Selected for retention in TIC by Cox regression analysis

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio; AUC = area under the ROC curve; p = p-value for the null hypothesis test that the true AUC = 0.50

The logistic regression model retained 4 of the 5 entered predictors: Age, BMI, Number of Dependents, and Years of Education. The model was statistically significant ($p < 0.001$) and yielded a Nagelkerke R^2 value of 0.134. Performance of the TIC with the 4 predictors retained by the logistic regression model is presented in Table 26. Calculations for TIC performance for all 4 tests positive were not completed because there were no male trainees with all 4 tests positive in the sample. Cross-tabulations for the first 3 levels of the TIC based on logistic regression are presented in Tables 27-29. Prior to providing data for the TIC, any male trainee had a pre-test probability of 15.6% for overuse injury based on the prevalence of overuse injuries among all male trainees. A male trainee with positive scores for any 3 of the 4 predictors in the TIC had a post-test probability of 90.6% for overuse injury.

Table 26. Performance of levels in the TIC based on logistic regression for prediction of overuse injuries for male trainees

	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability
Any 1 or more tests positive	0.58	0.30	1.89	0.61	15.6%	25.9%	10.1%
Any 2 or more tests positive	0.29	0.04	7.33	0.74	15.6%	57.5%	12.0%
Any 3 tests positive	0.08	1.00	51.96	0.92	15.6%	90.6%	14.5%
All 4 tests positive	*	*	*	*	*	*	*

*Not computed because no male trainees had all 4 tests positive

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio.

Table 27. Frequencies of male trainees who experienced vs. did not experience overuse injuries with any 1 or more tests positive vs. no tests positive in the TIC based on logistic regression

	Overuse Injury	No Overuse Injury
Any 1 or more tests positive	30	93
No tests positive	22	212

Percent correct classification: 67.8%

Table 28. Frequencies of male trainees who experienced vs. did not experience overuse injuries with any 2 or more tests positive vs. less than 2 tests positive in the TIC based on logistic regression

	Overuse Injury	No Overuse Injury
Any 2 or more tests positive	15	12
Less than 2 tests positive	37	293

Percent correct classification: 86.3%

Table 29. Frequencies of male trainees who experienced vs. did not experience overuse injuries with any 3 tests positive vs. less than 3 tests positive in the TIC based on logistic regression

	Overuse Injury	No Overuse Injury
Any 3 tests positive	4	0
Less than 3 tests positive	48	305

Percent correct classification: 86.4%

The Cox regression model retained 3 of the 5 entered predictors: Age, Number of Dependents, and Years of Education. The model was statistically significant ($p < 0.001$). Performance of the TIC for each of the 3 levels based on the Cox regression model is presented in Table 30. Calculations for TIC performance for all 3 tests positive were not completed because there were no male trainees with all 3 tests positive in the sample. Cross-tabulations for the first 2 levels of the TIC are presented in Tables 31-32. Prior to providing data for the TIC, any male trainee had a pre-test probability of 15.6% for overuse injury based on the prevalence of overuse injuries among all male trainees. A male trainee with positive scores for any 2 of the 3 predictors in the TIC had a post-test probability of 84.4% for overuse injury.

Table 30. Performance of levels in the TIC based on Cox regression for prediction of overuse injuries for male trainees

	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability
Any 1 or more tests positive	0.46	0.80	2.27	0.68	15.6%	29.6%	11.1%
Any 2 or more tests positive	0.19	0.99	29.33	0.81	15.6%	84.4%	12.3%
All 3 tests positive	*	*	*	*	*	*	*

*Not computed because no male trainees had all 3 tests positive

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio.

Table 31. Frequencies of male trainees who experienced vs. did not experience overuse injuries with any 1 or more tests positive vs. no tests positive in the TIC based on Cox regression

	Overuse Injury	No Overuse Injury
Any 1 or more tests positive	24	62
No tests positive	28	243

Percent correct classification: 74.8%

Table 32. Frequencies of male trainees who experienced vs. did not experience overuse injuries with any 2 or more tests positive vs. less than 2 tests positive in the TIC based on Cox regression

	Overuse Injury	No Overuse Injury
Any 2 tests positive	10	2
Less than 2 tests positive	42	303

Percent correct classification: 87.7%

ATTRITION: WOMEN

Among women in this study, 121 (29.1%) of 416 trainees were lost to attrition. For female trainees, 4 continuous-scale predictors (Table 33) and 1 categorical-scale predictor (Table 34) discriminated ($p \leq 0.05$) between those who completed BCT with their peers and those who did not. These predictor variables were Push-ups Initial Test, Sit-up Initial Test, Run Time Initial Test, Years of Education, and Pay Grade.

Table 33. Comparisons of means: attrition vs. completion of BCT for female trainees

Potential Predictor	Yes/No Completed BCT	N	Mean	Std. Deviation	t	df	p-value (2-tailed)
Push-ups Initial Test*	yes	294	10.16	8.94	2.99	402.00	0.003
	no	110	7.27	7.77			
Sit-ups Initial Test*	yes	294	24.66	9.08	2.05	402.00	0.041
	no	110	22.63	8.27			
Run Time Initial Test*	yes	293	10.28	1.45	-4.61†	155.26†	<0.001†
	no	107	11.20	1.87			
Age (years)	yes	295	21.89	4.14	0.56	414.00	0.577
	no	121	21.64	4.30			
Weight (lbs)	yes	295	138.78	20.68	-0.73	414.00	0.466
	no	121	140.44	22.04			
Height (in)	yes	295	64.38	2.51	-0.53	414.00	0.598
	no	121	64.53	2.68			
BMI (kg/m^2)	yes	295	23.47	2.87	-0.51	414.00	0.611
	no	121	23.63	3.00			
Number of Dependents	yes	212	0.43	0.86	0.66	306.00	0.509
	no	96	0.36	0.84			
Armed Forces Qualification Test	yes	212	55.13	18.80	1.00	306.00	0.319
	no	96	52.86	17.63			
Years of Education*	yes	212	12.44	1.17	2.27†	221.67†	0.024†
	no	96	12.16	0.96			
Average Household Income ($)	yes	207	53528.73	18949.33	0.59†	228.82†	0.554†
	no	93	52348.70	14408.72			

*Selected for entry into logistic regression analysis

†Adjusted for heterogeneity of variance

Table 34. Comparisons of frequencies: attrition vs. completion of BCT for female trainees (Chi-Square test results)

	Pearson Chi-Square	df	p-value (2-sided)
Pay Grade*	8.08	3	0.044
Race	1.70	5	0.889
Component	2.32	2	0.314
Fitness Assessment Program	5.13	2	0.077

*Selected for entry into logistic regression analysis

Table 35 contains derived cut score values, prognostic accuracy statistics, associated pre-test to post-test probability shifts, and results of ROC curve analyses for the 5 predictor variables selected for entry into the logistic regression model.

Table 35. Performance of selected single variables in prediction of attrition for female trainees

	Cut Score: positive test	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability	AUC	p
Push-ups Initial Test	<3	0.40	0.76	1.68	0.79	29.1%	40.8%	24.4%	0.603	0.032
Sit-ups Initial Test	<3	0.05	0.98	2.67	0.97	29.1%	52.3%	28.5%	0.577	0.030
Run Time Initial Test*	>13.96	0.05	0.99	6.07	0.96	29.1%	71.4%	28.2%	0.648	<0.001
Years of Education	<12	0.11	0.93	1.74	0.95	29.1%	41.6%	28.0%	0.569	0.054
Pay Grade	<E-2	0.62	0.49	1.25	0.75	29.1%	34.0%	23.6%	NA	NA

*Selected for retention in TIC by logistic regression analysis

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio; AUC = area under the ROC curve; p = p-value for the null hypothesis test that the true AUC = 0.50; NA = not applicable (categorical variables).

The logistic regression model retained only 1 of the 5 entered predictors: Run Time Initial Test. The model was statistically significant ($p < 0.001$) and yielded a Nagelkerke R^2 value of 0.087. Performance of the Run Time Initial Test as a univariate predictor of attrition for women is presented above in Table 35. A cross-tabulation of frequencies for the Run Time Initial Test is presented in Table 36. Prior to providing data for prediction, any female trainee had a pre-test probability of 29.1% for attrition based on the prevalence of attrition among all female trainees. A female trainee who could not complete the Run Time Initial Test in less than 13.96 minutes had a post-test probability of 71.7% for attrition.

Table 36. Frequencies of female trainees who completed vs. did not complete BCT with a positive test (>13.96 min) vs. a negative test (≤13.96 min) for the Run Time Initial Test

	Did Not Complete BCT	Completed BCT
>13.96 min	9	6
≤13.96 min	172	727

Percent correct classification: 80.5%

ATTRITION: MEN

Among men in this study, 76 (14.7%) of 518 trainees were lost to attrition. For male trainees, 1 continuous-scale predictor (Table 37) and no categorical-scale predictors (Table 38) discriminated ($p \leq 0.05$) between those who completed BCT with their peers and those who did not. The predictor was Push-ups Initial Test.

Table 37. Comparisons of means: attrition vs. completion of BCT for male trainees

Potential Predictor	Yes/No Completed BCT	N	Mean	Std. Deviation	t	df	p-value (2-tailed)
Push-ups Initial Test*	yes	441	29.66	11.00	2.86	90.31	0.005
	no	74	24.96	13.37			
Sit-ups Initial Test	yes	441	31.02	6.89	1.49	90.40	0.139
	no	74	29.49	8.36			
Run Time Initial Test	yes	440	8.45	1.50	-1.12	512.00	0.265
	no	74	8.66	1.54			
Age (years)	yes	442	21.98	3.93	0.04	516.00	0.969
	no	76	21.96	3.99			
Weight (lbs)	yes	442	173.53	29.80	0.16	516.00	0.874
	no	76	172.93	32.76			
Height (in)	yes	442	69.15	2.79	-1.26	516.00	0.209
	no	76	69.59	2.96			
BMI (kg/m^2)	yes	442	25.46	3.94	0.90	516.00	0.369
	no	76	25.01	4.14			
Number of Dependents	yes	307	0.37	0.82	-0.42	359.00	0.671
	no	54	0.43	0.79			
Armed Forces Qualification Test	yes	307	59.62	21.02	1.80	359.00	0.072
	no	54	54.13	18.20			
Years of Education	yes	307	12.48	1.38	1.79	107.82	0.076
	no	54	12.22	0.86			
Average Household Income ($)	yes	297	58512.42	27688.65	0.64	349.00	0.524
	no	54	55993.00	20430.89			

*Selected for entry into logistic regression analysis

‡Adjusted for heterogeneity of variance

Table 38. Comparisons of frequencies: attrition vs. completion of BCT for male trainees (Chi-Square test results)

	Pearson Chi-Square	df	p-value (2-sided)
Pay Grade	4.78	3	0.189
Race	11.29	6	0.080
Component	1.56	2	0.459
Fitness Assessment Program	0.66	2	0.718

Table 39 contains the derived cut score value, prognostic accuracy statistics, associated pre-test to post-test probability shifts, and results of the ROC curve analysis for the Push-ups Initial Test.

Table 39. Performance of the Push-ups Initial Test in prediction of attrition for male trainees

	Cut Score: positive test	Sn	Sp	PLR	NLR	Pre-test Probability	Positive Test Post-test Probability	Negative Test Post-test Probability	AUC	p
Push-ups Initial Test*	<11	0.18	0.96	4.84	0.86	14.7%	45.5%	12.8%	0.606	0.003

*Selected for retention in TIC by logistic regression analysis

Sn = sensitivity; Sp = specificity; PLR = positive likelihood ratio; NLR = negative likelihood ratio; AUC = area under the ROC curve; p = p-value for the null hypothesis test that the true AUC = 0.50

The logistic regression model retained the single entered predictor: Push-ups Initial Test. The model was statistically significant (p = 0.001) and yielded a Nagelkerke R^2 value of 0.037. Performance of the Push-ups Initial Test as a univariate predictor of attrition for men is presented above in Table 39. A cross-tabulation of frequencies for the Push-ups Initial Test is presented in Table 40. Prior to providing data for prediction, any male trainee had a pre-test probability of 14.7% for attrition based on the prevalence of attrition among all male trainees. A male trainee who performed less than 11 correct repetitions on the Push-ups Initial Test had a post-test probability of 45.5% for attrition.

Table 40. Frequencies of male trainees who completed vs. did not complete BCT with a positive test (<11 repetitions) vs. a negative test (≥ 11 repetitions) for the Push-ups Initial Test

	Did Not Complete BCT	Completed BCT
<11 push-ups	13	16
≥ 11 push-ups	61	425

Percent correct classification: 85.1%

DISCUSSION

The analytic procedure employed in this study permitted identification of multivariate models for predicting APFT failure in both men and women, and for predicting overuse injuries in men during BCT. The process yielded a single best predictor for attrition for both men and women, and for overuse injuries in women.

The PLRs (22.8, 10.5) associated with TICs for predicting failure of the APFT in both genders and for predicting overuse injuries for male trainees (PLR = 51.1 with logistic regression; 29.3 with Cox regression) met or exceeded the criterion of Jaeschke et al.[28] for large and often "conclusive" shifts from pre-test probability to post-test probability for both men and women. The ability to predict during the first week of BCT that an individual has a probability of 84% for failing the APFT or a 91% probability of an incurring an overuse injury would be extremely valuable to commanders. Such a characterization of individual trainees with high levels of certainty would permit classification of trainees according to risk and might facilitate

tailoring of the training experience to mitigate risk. However, these preliminary predictive estimates are based only on this single derivation study using a single data set, and cannot be considered appropriate for field use. Useful predictive models may be developed with additional prospective studies designed to capture all relevant predictors[20] followed by additional validation studies using independent data sets.

Although the PLRs (1.3, 1.5) for single predictors of overuse injuries in women was disappointingly small using either regression method, PLRs (6.1, 4.8) for tests predicting attrition in both genders achieved or approached the magnitude required for moderate shifts in post-test probability. These probability shifts from 29% to 72% for women, and from 15% to 46% for men could be of potential importance in identifying individuals at risk for attrition. It is doubtful that shifting probability of overuse injuries in women from 38% to 44% represents any helpful reduction in uncertainty.

Overall results using Cox regression were similar to those using logistic regression in TIC derivation for overuse injury prognosis. Although the two methods selected different single predictors for women, the post-test probability estimates were similar (30.7% vs 34.7%). The two methods selected the same 3 predictors for men, although logistic regression also selected one additional predictor. Consequently, post-test probability estimates for the two TICs predicting injury for men were not very different (90.6% vs. 84.4%). These results do not rule out the possibility that more inclusive TICs based on prospective studies with more variables for analysis might achieve better results using Cox regression for outcomes with unequal time at risk for subjects who attrite.

Concato et al.[35] recommended that a minimum of 10 “events” (subjects with negative outcomes of interest) be present for every predictor in a multivariate regression model in order to avoid the problems of “overfitting” and questionable accuracy in the models. Numbers of trainees with negative training outcomes were well over this minimum criterion for all of the models derived in this study.

The TICs for predicting APFT failure for both genders included baseline tests that were very similar to elements of the APFT (1-minute push-ups,1-minute sit-ups tests, and 1-mile run). Although independence of diagnostic or predictive tests from criterion reference tests is important in any study of diagnosis, prognosis, or prediction, the temporal separation of 6 weeks between the initial fitness assessment and the APFT, and the time and distance differences, helped reduce the possibility of inflating the estimates of predictive power.

It was interesting that cut scores derived from the ROC curve analysis approximated minimum performance standards in place for determining FAP status. For example, the cut scores of 2 push-ups, 10 sit-ups, and 10.7 minutes to run for the Initial Test as predictors of APFT failure for women are reasonably close to the fitness criteria of 3 push-ups, 17 sit-ups, and 10.5 minutes to determine FAP status. Similarly, the fitness criteria of 13 push-ups, 17 sit-ups, and 8.5 minutes to determine

FAP status for men are fairly close to the cut scores to predict APFT failure for men: 13 push-ups, 21 sit-ups, and 9.4 minutes to run for the Initial Test.

Fitness Assessment variables (push-ups, sit-ups and 1-mile run) were significant univariate predictors of all 3 outcome variables for both men and women; they were included in the multivariate prediction models of all but one outcome (overuse injury for men). It is intuitively appealing that initial fitness should have some predictive role in determining the ability to pass the APFT. However, both initial training status and genetic endowment play a role in the adaptive response to a physical training program. Individuals engaged in exercise programs of virtually identical frequency, intensity, and duration show great variations in improvements in aerobic power, endurance performance, and anaerobic capacity (24, 25, 26, 27, 28). On the other hand, the role of fitness in predicting attrition is not as apparent but may be related to BCT task performance. On the aerobic and muscular endurance tasks performed in BCT (e.g., running, road marching, obstacle course, bayonet course, etc), less fit trainees will perform at a higher percentage of their maximal physical capacity. They will perceive BCT tasks as being more difficult (29, 30) and they will fatigue more rapidly (31,32,33). These factors influence their ability to train, may influence their motivation, and may influence how their drill sergeants and peers view their performance. A previous study has emphasized the importance of physical fitness as a risk factor for discharge (34).

Results of this study were greatly affected by the decision to select cut scores for continuous predictor variables that would yield the highest possible values for specificity and PLRs. This choice reflects a philosophy that it is more important to identify a trainee who has high risk for injury or failure than to identify a trainee who has low risk. However, both mistakes in prediction have negative consequences. An ideal predictive test would simultaneously minimize false negative and false positive test results, but this ideal circumstance is rarely encountered when trying to predict complex phenomena. There can be little doubt that negative outcomes in BCT are multifactorial; therefore, a multivariate approach to prediction such as TIC development has intuitive appeal. Still, as demonstrated in the results of this study, multivariate TICs still tend to be less than ideal and tend to force choices between selecting for either greater PLRs or greater NLRs, but not both. The practical choice between avoiding false negative results and avoiding false positive results is also multifactorial, and will be affected by economic, political, and ethical considerations. Analysis of other potentially predictive factors known to effect the outcome variables could improve predictive power (e.g., for injury, cigarette smoking and pre-training physical activity; for attrition, race, and prior job history). However, it is highly unlikely that even an exhaustive set of predictors for analysis could yield models with zero false positives and zero false negatives. Decisions made on the basis of imperfect predictive models have potential to disadvantage individuals who might be incorrectly classified by the models. Furthermore, if models should be used to disqualify individuals from training, institutional efforts to produce trained Soliders could be disadvantaged to the extent that false positive predictions and false negative predictions are inherent in the models.

This study was not optimized for derivation of the best available TICs because we used an existing data set with only a limited number of potential predictor variables. However, results from this study demonstrate the potential usefulness of the analytic method used in TIC derivation, and the potentially powerful shifts in pre-test to post-test probabilities that can result from multivariate models compared to using single predictors in isolation.

CONCLUSIONS

Large and potentially conclusive shifts in pre-test to post-test probability were observed with TICs derived to predict APFT failure for both men and women, and to predict overuse injuries in men. These multivariate models suggest that negative training outcome probabilities as high as 91% might be estimated for individual trainees, given positive results for predictive test clusters. Moderate probability shifts were seen with the single tests identified to predict BCT attrition for both men and women. No useful model for predicting overuse injuries in women was derived from the methods employed in this study.

RECOMMENDATIONS

Future studies with the goal of TIC development should be planned in sequence. First, prospective studies should collect data on the broadest possible spectrum of known and suspected risk factors for negative training outcomes in order to derive more robust and inclusive TICs. Second, TICs must be validated on a second, independent sample of trainees before recommendation for use. Third, research is needed to study the impact of implementing validated TICs on cost, outcome, and behavior.

REFERENCES

1. Knapik JJ, Darakjy S, Scott S, et al. *Evaluation of Two Army Fitness Programs: The TRADOC Standardized Physical Training Program for Basic Combat Training and the Fitness Assessment Program.* Aberdeen Proving Ground, MD: U.S. Army Center for Health Promotion and Preventive Medicine; 2004. 12-HF-5772B-04.

2. Jones BH, Bovee MW, Harris JM, 3rd, Cowan DN. Intrinsic risk factors for exercise-related injuries among male and female army trainees. *Am J Sports Med.* Sep-Oct 1993;21(5):705-710.

3. Jones BH, Cowan DN, Tomlinson JP, Robinson JR, Polly DW, Frykman PN. Epidemiology of injuries associated with physical training among young men in the army. *Med Sci Sports Exerc.* Feb 1993;25(2):197-203.

4. Jones BH, Cowan DN, Knapik JJ. Exercise, training and injuries. *Sports Med.* Sep 1994;18(3):202-214.

5. Knapik JJ, Sharp MA, Canham-Chervak M, Hauret K, Patton JF, Jones BH. Risk factors for training-related injuries among men and women in basic combat training. *Med Sci Sports Exerc.* Jun 2001;33(6):946-954.

6. Knapik JJ, Canham-Chervak M, Hauret K, et al. Seasonal variations in injury rates during US Army Basic Combat Training. *Ann Occup Hyg.* Jan 2002;46(1):15-23.

7. Knapik JJ, B.H.Jones, Hauret KG, Darakjy S and Piskator G. A review of the literature on attrition from the military services: risk factors and strategies to reduce attrition. Aberdeen Proving Ground, MD: US Army Center for Health Promotion and Preventive Medicine. Technical Report No. 12-HF-01Q9A-04, 2004.

8. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA.* Feb 12 1997;277(6):488-494.

9. McGinn T, Guyatt G, Yyer P, C.D. N, Stiell I. Diagnosis: Clinical Prediction Rules. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature- A Manual for Evidence-Based Clinical Practice*. Chicago: AMA Press; 2002:471-483.

10. Wainner RS, Fritz JM, Irrgang JJ, Boninger ML, Delitto A, Allison S. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine*. Jan 1 2003;28(1):52-62.

11. Wainner RS, Fritz JM, Irrgang JJ, Delitto A, Allison S, Boninger ML. Development of a clinical prediction rule for the diagnosis of carpal tunnel syndrome. *Arch Phys Med Rehabil.* 2005 Apr;86(4):609-18.

12. Ebell MH, Smith MA, Barry HC, Ives K, Carey M. The rational clinical examination. Does this patient have strep throat? *JAMA*. 2000 Dec 13;284(22):2912-8.

13. Ferreira G, Carson JL. Clinical prediction rules for the diagnosis of pulmonary embolism. *Am J Med*. 2002 Sep;113(4):337-8.

14. Morise AP, Haddad WJ, Beckner D. Development and validation of a clinical score to estimate the probability of coronary artery disease in men and women presenting with suspected coronary disease. *Am J Med*. 1997 Apr;102(4):350-6.

15. Wells PS, Hirsh J, Anderson DR, et al. Accuracy of clinical assessment of deep-vein thrombosis. *Lancet*. 1995;345:1326–1330.

16. Flynn T, Fritz J, Whitman J, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine.* Dec 15 2002;27(24):2835-2843.

17. Sutlive TG, Mitchell SD, Maxfield SN, et al. Identification of individuals with patellofemoral pain whose symptoms improved after a combined program of foot orthosis use and modified activity: a preliminary investigation. *Phys Ther.* Jan 2004;84(1):49-61.

18. Solomon DH, Avorn J, Warsi A, et al. Which patients with knee problems are likely to benefit from nonarthroplasty surgery? Development of a clinical prediction rule. *Arch Intern Med.* Mar 8 2004;164(5):509-513.

19. Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil.* 2005 Sep;86(9):1753-62.

20. Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapist practice. *Phys Ther.* 2006 Jan;86(1):122-31.

21. Schatz M, Cook EF, Joshua A, Petitti D. Risk factors for asthma hospitalizations in a managed care organization: development of a clinical prediction rule. *Am J Manag Care.* Aug 2003;9(8):538-547.

22. Selby JV, Karter AJ, Ackerson LM, Ferrara A, Liu J. Developing a prediction rule from automated clinical databases to identify high-risk patients in a large population with diabetes. *Diabetes Care.* Sep 2001;24(9):1547-1555.

23. van Walraven C, Hart RG, Wells GA, et al. A clinical prediction rule to identify patients with atrial fibrillation and a low risk for stroke while taking aspirin. *Arch Intern Med.* Apr 28 2003;163(8):936-943.

24. Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? *Aust J Physiother*. 2006;52(3):157-63.

25. Kuijpers T, van der Windt DA, Boeke AJ, Twisk JW, Vergouwe Y, Bouter LM, van der Heijden GJ. Clinical prediction rules for the prognosis of shoulder pain in general practice. Pain. 2006 Feb;120(3):276-85.

26. Dujardin B, Van den Ende J, Van Gompel A, Unger JP, Van der Stuyft P. Likelihood ratios: a real improvement for clinical decision making? *Eur J Epidemiol.* Feb 1994;10(1):29-36.

27. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA.* Feb 2 1994;271(5):389-391.

28. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* Mar 2 1994;271(9):703-707.

29. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5):257

30. Physical Fitness Training. Army Field Manual (FM) 21-20. Washington DC: Headquarters, Department of the Army, 1992.

31. Kilburn MR, Hanser LM, Klerman JA. The AFQT and Its Role in Enlistment. In *Estimating AFQT Scores for National Educational Longitudinal Study (NELS) Respondents*. Santa Monica, CA: Rand Corporation; 1998:5-8.

32. Stevens J. *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986: 58.

33. Altman DG, Machin D, Bryand TN, Gardner MJ. *Statistics with Confidence*. 2nd ed. Bristol, U.K.: BMJ Books; 2000: 110.

34. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther.* Sep 2001;81(9):1546-1564.

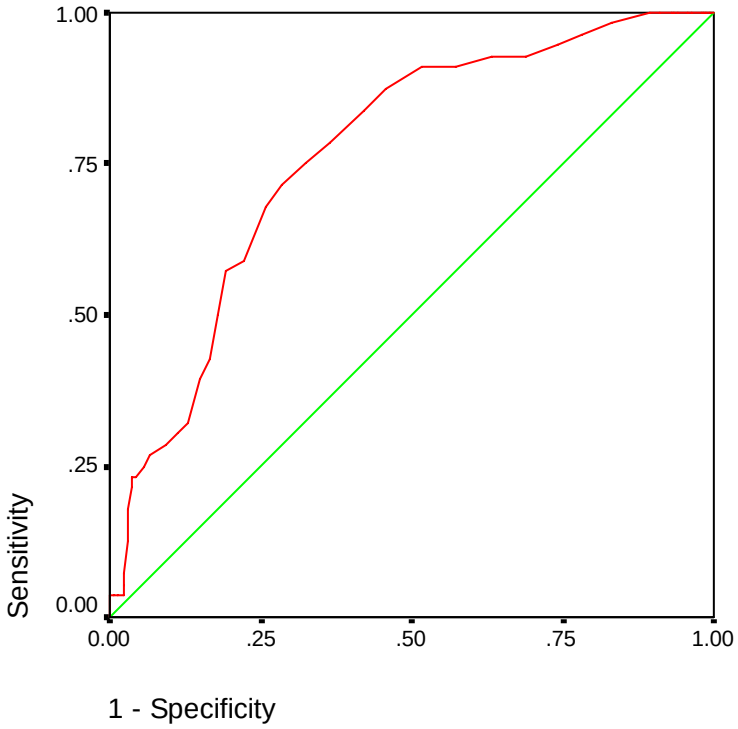
35. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Ann Intern Med. 1993 Feb 1;118(3):201-10.

APPENDIX A: RECEIVER-OPERATOR CHARACTERISTIC CURVES

Receiver-operator characteristic curves are presented below for continuous-scale predictors entered into logistic regression analysis for each of the 6 test item clusters.

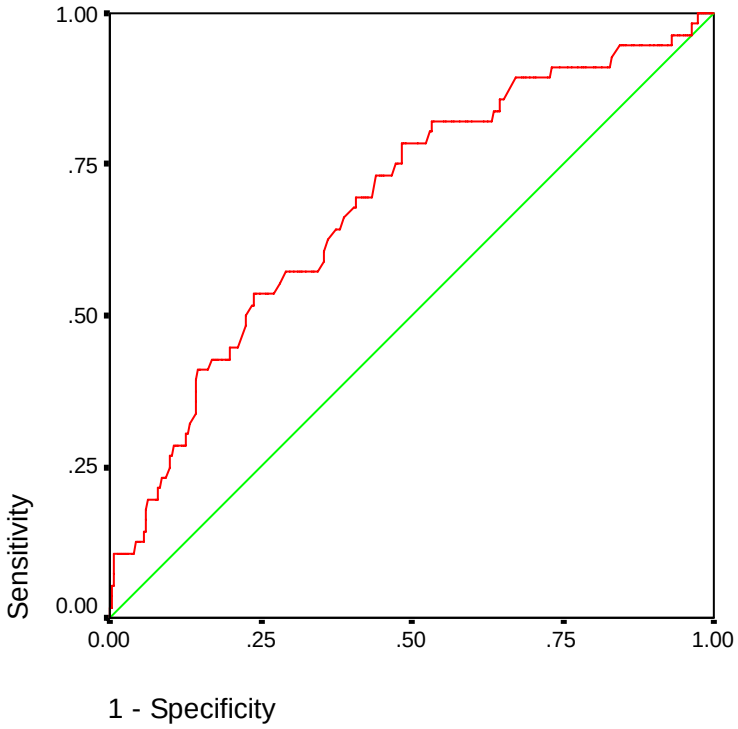
APFT FAILURE: WOMEN

Figure 2. ROC Curve for prediction of APFT failure: female Sit-ups Initial Test



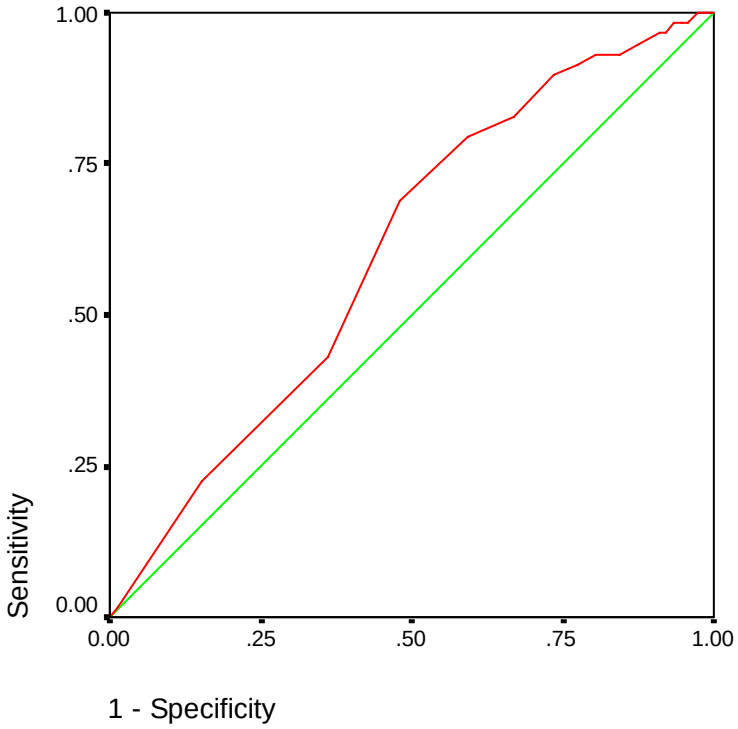
AUC = 0.767, p < 0.001

Figure 3. ROC Curve for prediction of APFT failure: female Run Time Initial Test



AUC = 0.685, p < 0.001

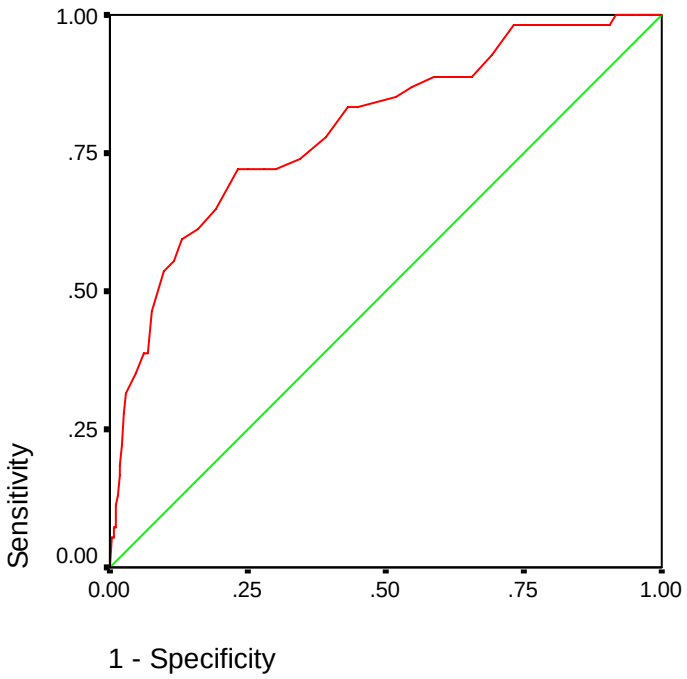
Figure 4. ROC Curve for prediction of APFT failure: female Age



AUC = 0.606, p = 0.012

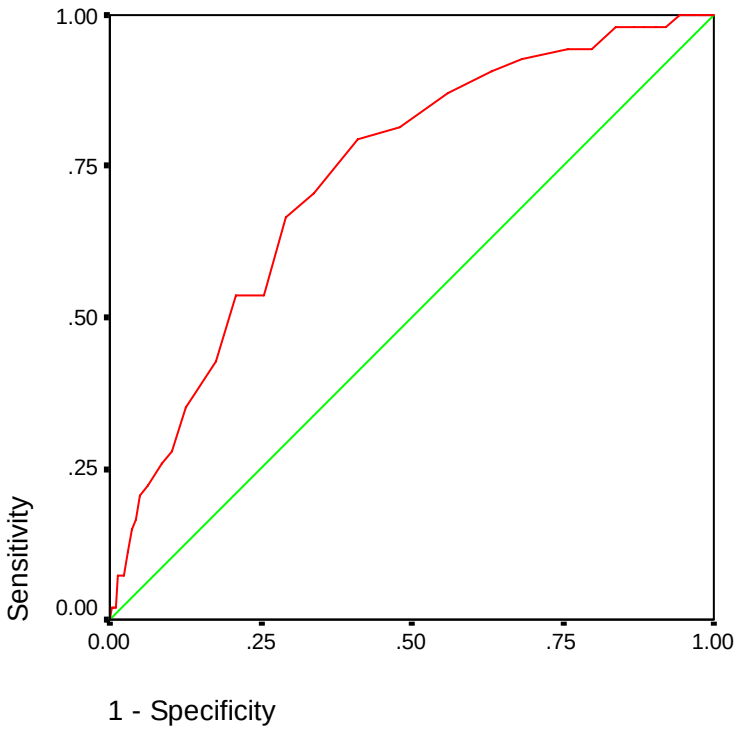
APFT FAILURE: MEN

Figure 5. ROC Curve for prediction of APFT failure: male Push-ups Initial Test



AUC = 0.795, p < 0.001

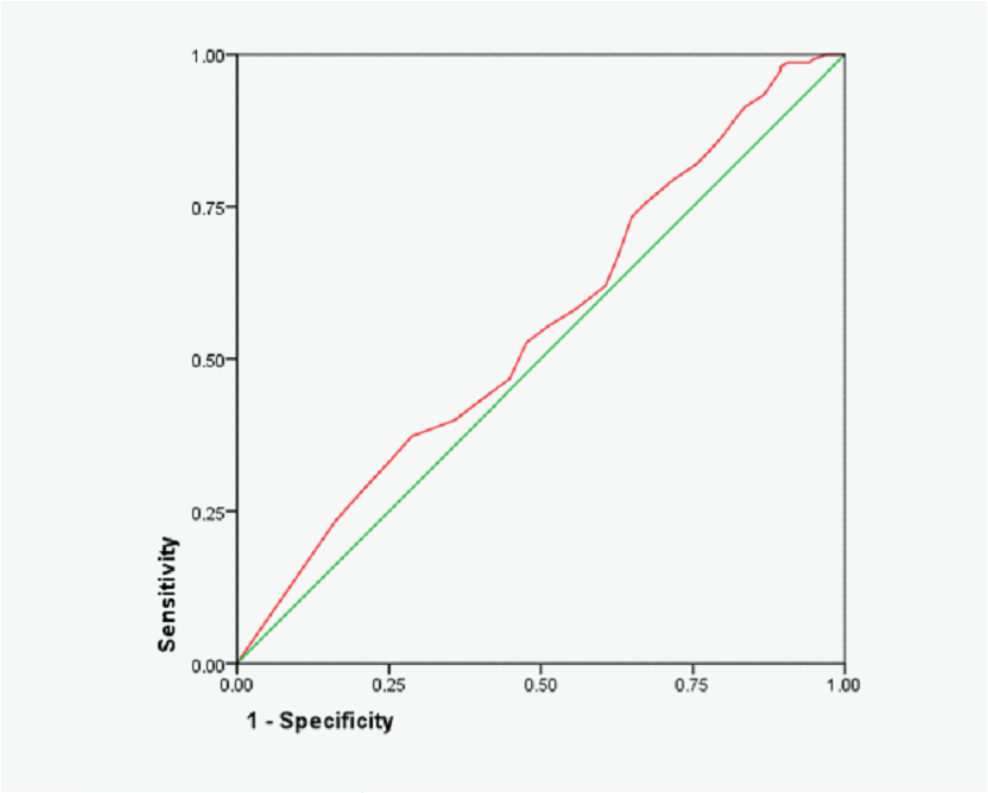
Figure 6. ROC Curve for prediction of APFT failure: male Sit-ups Initial Test



AUC = 0.734, p < 0.001

OVERUSE INJURIES: WOMEN

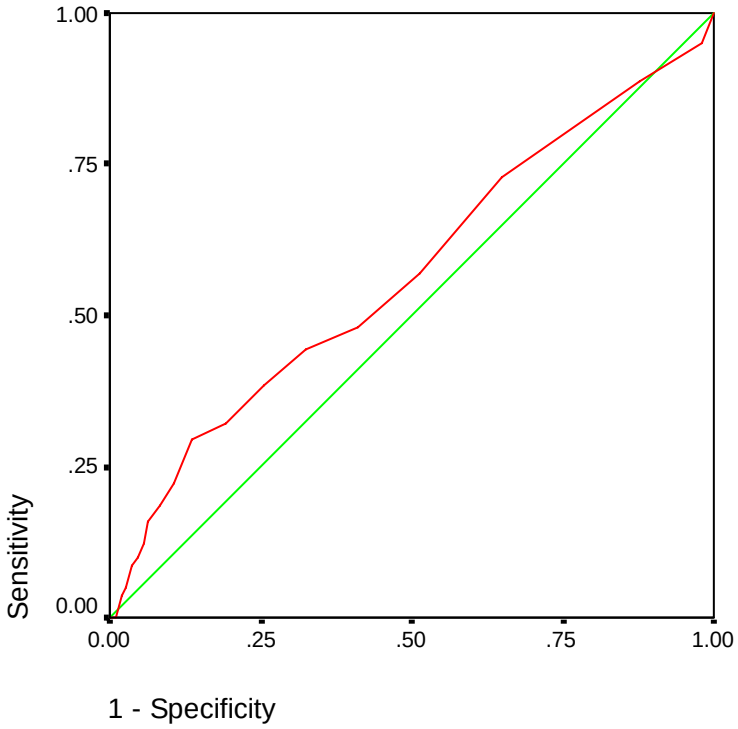
Figure 7. ROC Curve for prediction of overuse injuries: female Push-ups Initial Test



AUC = 0.553, p = 0.076

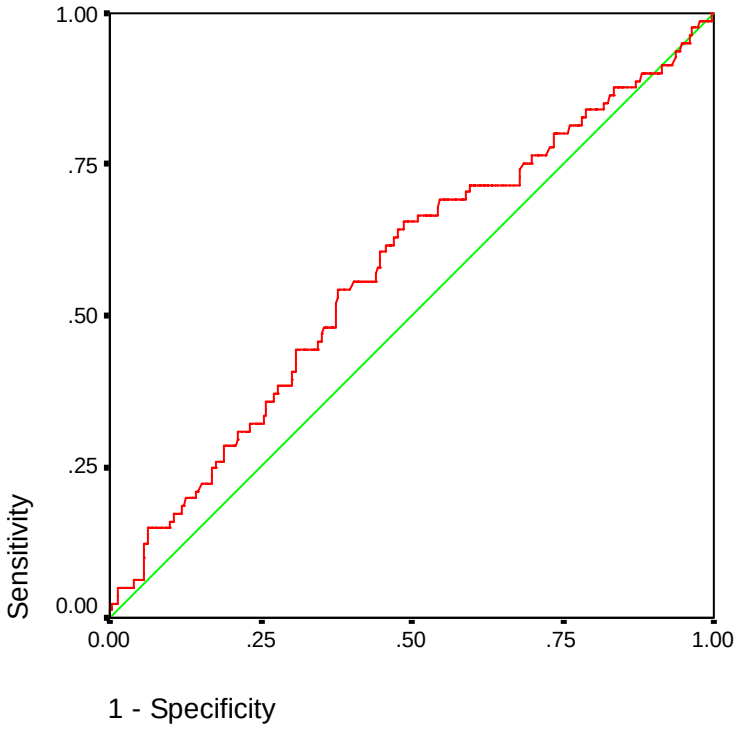
OVERUSE INJURIES: MEN

Figure 8. ROC Curve for prediction of overuse injuries: male Age



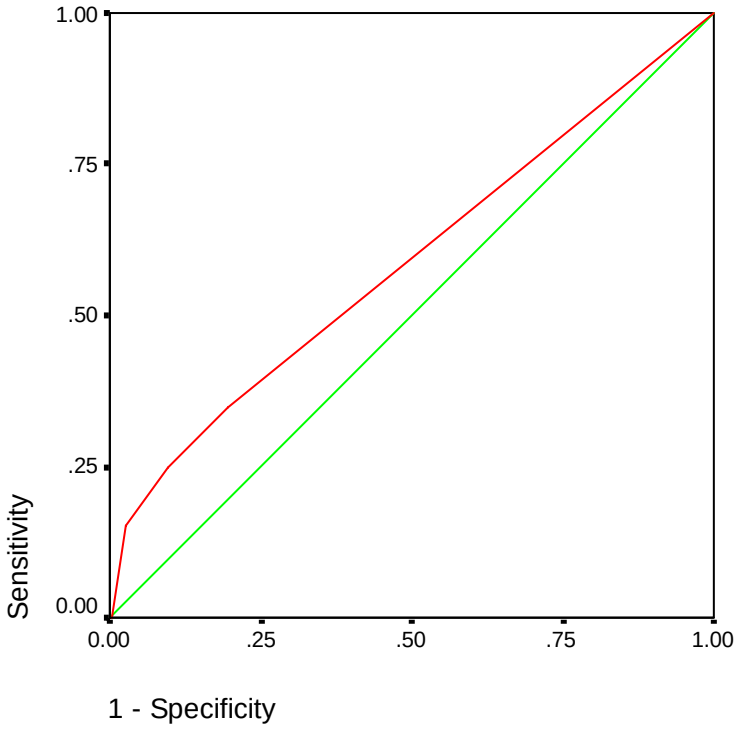
AUC = 0.569, p = 0.050

Figure 9. ROC Curve for prediction of overuse injuries: male Body Mass Index



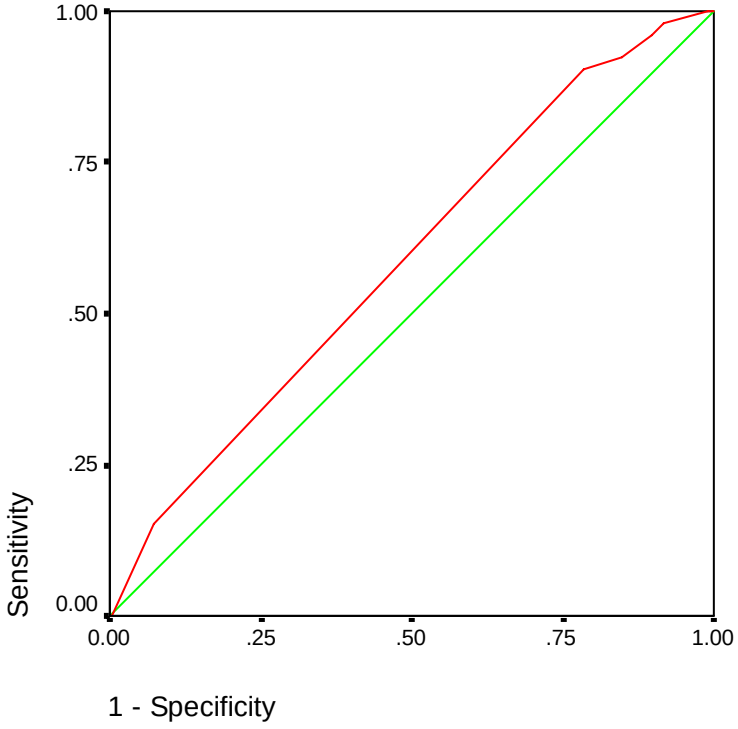
AUC = 0.574, p = 0.034

Figure 10. ROC Curve for prediction of overuse injuries: male Number of Dependents



AUC = 0.587, p = 0.044

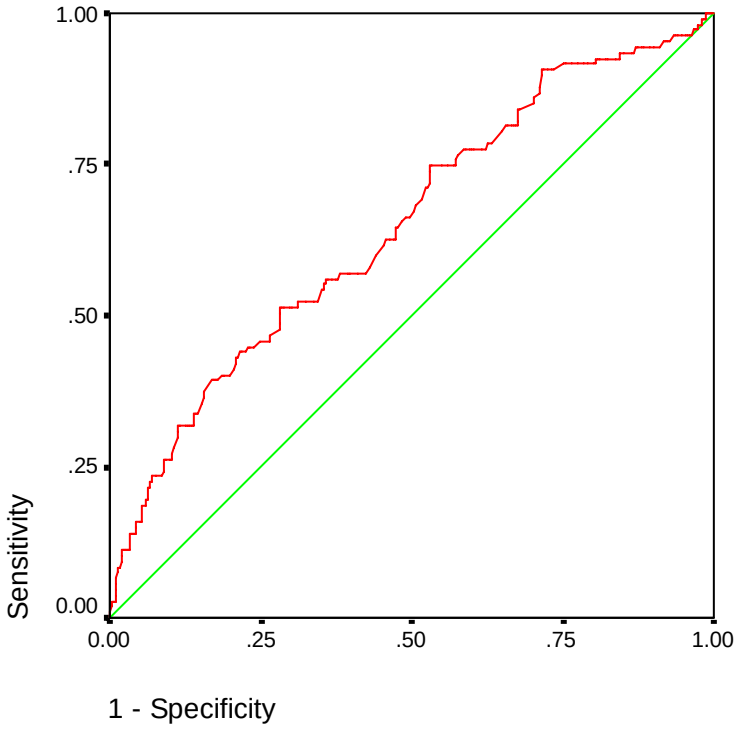
Figure 11. ROC Curve for prediction of overuse injuries: male Years of Education



AUC = 0.588, p = 0.043

ATTRITION: WOMEN

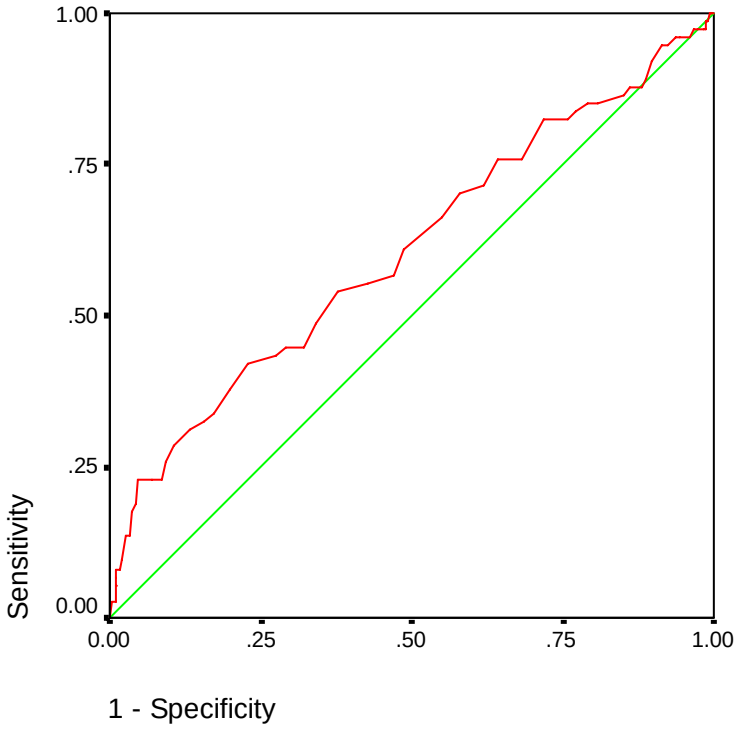
Figure 12. ROC Curve for prediction of attrition: female Run Time Initial Test



AUC = 0.648, p < 0.001

ATTRITION: MEN

Figure 13. ROC Curve for prediction of attrition: male Push-ups Initial Test



AUC = 0.606, p = 0.003

APPENDIX B: COORDINATE POINTS FOR THE ROC CURVE

The ROC curve coordinate points presented below illustrate the method of selecting a cut score with a specific test performance goal in mind. In this case the goal was to minimize the number of false positive tests by selecting a cut score yielding a high specificity (low value for 1-specificity) and a high positive likelihood ratio. Selection of a cut score is a judgment process accomplished by viewing the plotted figure (see Appendix A, Figure 4) for overall characteristics of the scale, then scanning the coordinate plots. Ideally, one would select a cut score that maximizes sensitivity and specificity simultaneously. Commonly, a choice is made to maximize PLR or NLR. However, the choice of a cut score may not yield the absolute highest PLR or lowest NLR due to considerations for optimizing multiple attributes simultaneously.

Table 41. Coordinate points from the ROC curve analysis using Push-ups Initial Test to predict APFT failure for male trainees

Positive if Less Than or Equal To*	Sensitivity	1 - Specificity	PLR	NLR
-1	0	0		
0.5	0.056	0.002	22.444	0.947
1.5	0.056	0.007	7.481	0.952
2.5	0.074	0.007	9.975	0.933
4.5	0.074	0.01	7.481	0.935
6.5	0.111	0.01	11.222	0.898
7.5	0.13	0.015	8.728	0.883
8.5	0.167	0.017	9.619	0.848
9.5	0.185	0.02	9.352	0.831
10.5	0.222	0.022	9.975	0.795
11.5	0.278	0.025	11.222	0.741
12.5**	**0.315**	**0.03**	**10.599**	**0.706**
13.5	0.352	0.047	7.481	0.68
14.5	0.389	0.062	6.284	0.651
15.5	0.389	0.069	5.611	0.657
16.5	0.463	0.077	6.033	0.582
17.5	0.537	0.097	5.563	0.512
18.5	0.556	0.116	4.775	0.503
19.5	0.593	0.131	4.517	0.469
20.5	0.611	0.161	3.798	0.463
21.5	0.648	0.191	3.401	0.435
22.5	0.722	0.23	3.137	0.361
23.5	0.722	0.25	2.889	0.37
24.5	0.722	0.277	2.605	0.384
25.5	0.722	0.3	2.411	0.397
26.5	0.741	0.344	2.153	0.395
27.5	0.778	0.391	1.989	0.365
28.5	0.833	0.431	1.935	0.293

29.5	0.833	0.45	1.85	0.303
30.5	0.852	0.517	1.647	0.307
31.5	0.87	0.547	1.591	0.286
32.5	0.889	0.587	1.515	0.269
33.5	0.889	0.614	1.448	0.288
34.5	0.889	0.656	1.355	0.323
35.5	0.926	0.693	1.336	0.241
36.5	0.981	0.73	1.344	0.069
37.5	0.981	0.743	1.322	0.072
38.5	0.981	0.765	1.283	0.079
39.5	0.981	0.785	1.251	0.086
40.5	0.981	0.829	1.184	0.108
41.5	0.981	0.842	1.166	0.117
42.5	0.981	0.861	1.139	0.134
43.5	0.981	0.869	1.13	0.141
44.5	0.981	0.886	1.108	0.163
45.5	0.981	0.906	1.083	0.197
46.5	1	0.916	1.092	0
47.5	1	0.928	1.077	0
48.5	1	0.936	1.069	0
49.5	1	0.953	1.049	0
50.5	1	0.965	1.036	0
51.5	1	0.973	1.028	0
52.5	1	0.98	1.02	0
54.5	1	0.985	1.015	0
56.5	1	0.988	1.013	0
57.5	1	0.993	1.007	0
60.5	1	0.998	1.002	0
64	1	1	1	

*Units are repetitions of push-ups; raw scores are integers

**Selected as cut score

PLR = positive likelihood ratio; NLR = negative likelihood ratio.

The Publishing Division
of the Massachusetts Medical Society

Publishers of
The New England Journal of Medicine, Journal Watch Newsletters, & AIDS Clinical Care

MMS Reference Number: PS - 2007 - 2316
MMS Invoice Number: RY - 2007 - 2316

Grant of Permission

January 05, 2007

US Army - USARIEM
Dr. Stephen Allison
Military Performance Division
42 Kansas Street
Natick, MA 01760-5007

Dear Dr. Allison,

Thank you for your interest in our copyrighted material, and for requesting permission for its use.

Permission is granted for limited, non-exclusive educational use of the material requested, subject to all the terms and conditions outlined throughout this document. Please review all of the following pages, including the "Basic Provisions of Grant of Permission" as well as "Items Covered by Grant of Permission." A Permissions Invoice is included as the last page, if applicable.

Thank you for your patience while your request was being processed. If you wish to contact us further, please use the address below, and cite our reference numbers on any correspondence.

Sincerely,

Jennifer Moran

Jennifer L. Moran

Sr. Rights & Permissions Representative

Publishing Division of the Massachusetts Medical Society
Department of Permissions & Licensing
860 Winter Street, Waltham, Massachusetts 02451-1413 USA
Tel: (781) 434 7382 · Fax: (781) 434 7633 · permissions@nejm.org

Items Covered by Grant of Permission

The Publishing Division
of the Massachusetts Medical Society

Department of Permissions & Licensing
860 Winter Street, Waltham, Massachusetts 02451-1413 USA
Telephone: (781) 434-7382 fax: (781) 434-7633

MMS Reference Number: PS - 2007 - 2316
MMS Invoice Number: RY - 2007 - 2316

Source Information							Further Conditions				
Source: The New England Journal of Medicine											
Volume	Pages	Pub. Date	Author(s)	Article Title	Type	Item	Format	Language	Adapted	Dollar Amount	Customer Reference
293	257- 257	07/31/1975	Terrence J. Fagan, M.D.	Nomogram for Bayes's Theorem	F	F1	On-line & Print	English	Y	0.00	Fig 1

The following information has been provided for us in your letter of request.

Year of Publication: 2007
Publisher: US Army
Title of Work: US Army Technical Report
Chapter/Article Title: Preliminary Derivation of Test Item Clusters for Predicting Injuries...
Author: Allison SC, Knapik JJ, Sharp MA
Editor:
Edition: